Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Contrast Effects in the Verbal Conditioning of Meaning: F. J. DI VESTA 535
The Spacing of Sequentially Dependent Trials in Probability Learning: J. P. CHAPMAN 545
Yoked Comparisons of Instrumental and Classical Eyelid Conditioning: J. W. MOORE AND I. GORMEZANO 552
The Intertrial Interval in Classical Conditioning: W. F. PROKASY AND F. L. WHALEY 560
Mediation in Paired-Associate Learning: N. E. MCGEHEE AND R. W. SCHULZ 565
Studies of Distributed Practice: XXI. Effect of Interference from Language Habits: B. J. UNDERWOOD AND R. W. SCHULZ 571
Semantic Satiation among Bilinguals: L. A. JAKOBOVITS AND W. E. LAMBERT 576
Response Pretraining: Differentiation or Availability?: E. SALTZ 583
The Effect of Stress on Stimulus Differentiation: E. SALTZ AND W. J. RIACH 588
Familiarity and Apparent Complexity of Random Shapes: A. G. GOLDSTEIN 594
Memory Effects in Concept Learning: E. B. HUNT 598
Postrest Motor Learning Performance as a Function of Degree of Learning: J. C. JAHNKE .. 605
The CS-UCS Interval in Conditioning Short- and Long-Latency Responses: J. E. JONES 612
The Retention of Individual Items: B. B. MURDOCK, JR. 618
Supplementary Report: Perception of Stimuli of Varying Dimensionality: P. G. SHINKMAN .. 626
Supplementary Report: A Selection Artifact in Rock's Study of the Role of Repetition: J. P. WILLIAMS 627
Supplementary Report: Alternation in a T Maze with Habit Held Constant: M. E. THOMPSON AND E. J. MARTIN 629
Supplementary Report: Unlearning the Gambler's Fallacy: H. LINDMAN AND W. EDWARDS .. 630
Supplementary Report: Two-Choice Decision Behavior with Many Alternative Events: R. A. GARDNER AND J. B. FORSYTHE 631
Supplementary Report: Short-Term Retention as a Function of Average Storage Load: K. E. LLOYD 632

American Psychological Association

Vol. 62, No. 6 December 1961

Important Editorial Announcements on Cover 4

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*. When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1961 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN
Production Manager

HELEN ORR
Promotion Manager

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1962 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 62, NO. 6 DECEMBER 1961

CONTRAST EFFECTS IN THE VERBAL CONDITIONING OF MEANING[1]

FRANCIS J. DI VESTA

Syracuse University

The present study extends the findings of earlier investigations of contrast and assimilation effects to attitude change accomplished through semantic mediation processes. The rationale is based principally on the concept-formation approach to attitude acquisition (Rhine, 1958). A basic assumption in this orientation is that a concept is defined by some principle or common dimension that permits the classification of objects. The principle is abstracted as a feature of a class of stimulus patterns through recurrent experiences. Such classifications can be made on the basis of both the objective and evaluative characteristics of the dimension of meaning. Thus, in addition to whatever objective characteristics contribute to the formation of a concept, it is conceivable that pleasant experiences might add the dimension of *good* to the meaning of a concept while unpleasant experiences or adverse information from a trustworthy source might add the dimension of *bad* to the concept. For all objects and experiences an essential condition for concept formation is the associations between a common response and a variety of stimuli.

Among the complex behavioral processes, the naming response is a most important basis for a common response. A simple example is that a common name can be attached to a series of objects, all of which have different physical characteristics to form a *class*, for example, the word *furniture* may be applied to many different objects which are physically quite dissimilar. The name, or verbal response, provides the basis for a common mediating process. As mediators they produce stimuli which may then become conditioned to other mediating, instrumental, or attitudinal responses, thereby permitting the facilitation of new learning whether cognitive or affective. Demonstration of the feasibility for producing attitudinal change within this theoretical framework has been provided through the use of what Staats and Staats (1957, 1958) refer to as "classical

[1] This research was supported by Research Grant M-2900 from the National Institute of Mental Health, United States Public Health Service. Thanks are due to Marilyn Rothschild and Fred W. Naugle for their invaluable assistance in collecting the present data.

conditioning." The effects of differential reinforcement on attitude strength, development, and change has been demonstrated by Rhine and Silun (1958). That evaluative stimulus components comprise one dimension of meaning which mediate verbal responses in forming associations is documented by the factorial analyses of meaning (Osgood, Suci, & Tannenbaum, 1957).

The research designs used in the studies exemplified by those of Staats and Staats (1957), as well as that of Rhine and Silun (1958), typically followed the pattern of developing two attitudes concurrently. For example, one concept-attitude was developed by pairing one name with words having positive-evaluative meanings and a second concept-attitude was developed by pairing another name with words having negative-evaluative meanings; or, one of the aforementioned was used in combination with the development of a "neutral" attitude. Treatment effects were determined by comparing the intra-individual changes or difference between the two attitudes following the experimental treatments.

In none of these studies was the possibility of a direct-opposite effect considered as an alternative for the formation of one of the concepts. In view of much of the recent (as well as the classical) literature on contrast effects we would hypothesize that similar effects will occur in the development of attitudes by conditioning. Changes are to be anticipated under these circumstances, in attitudes concerning related issues as well as the specific target issue toward which the persuasive effort is directed. In similar fashion, the literature in which a contrast effect has been demonstrated in judgments of simple stimulus materials (Harvey & Caldwell, 1959; Sherif, Taub, & Hovland, 1958) suggests the occurrence of a similar phenomenon in the simultaneous development of two attitudes. Under these circumstances, the more intense of the two stimuli serves as an anchor or reference point against which the other stimulus (attitude object) is evaluated. While the changes may either accompany the change in the target attitude or may occur instead of it, it is the former condition that is of present concern.

Thus, if contrast effects are to be demonstrated in the evaluation of a poorly structured concept, polarization will occur toward a position opposite to that evaluation made of the anchoring (target) concept-attitude. The degree of polarization thus will be inversely related to the scale position of the anchor object. The more remote the anchors the greater will be the displacement of the scale value of the related attitude from that of the anchor.

The general hypothesis is that a contrast effect is obtained when two attitudes are formed concurrently through verbal conditioning. The more specific hypothesis is that if the target concept-attitude is formed on the basis of a conceptual dimension such as evaluation, potency, or activity, the change in the *accompanying* (neutral) concept will also be on that dimension but toward the opposite pole. Since the concept dimensions of potency and activity also have high factorial loadings on the evaluative dimensions, it is assumed that a change in one of these concept dimensions will also be accompanied by a change in the evaluative dimension.

These general assumptions are suggested in experiments where judgments have been made on simple stimulus materials (Sherif et al., 1958), and in experiments where both source and content were studied

jointly (Hovland, Harvey, & Sherif, 1957) as variables influencing reactions to persuasion. None of the aforementioned studies has been concerned with contrast effects in laboratory situations where complex verbal stimuli were used in the development of attitudes, or where the development of one attitude was modified by the development of the other, while simultaneously minimizing or eliminating the complicating effects of source.

METHOD

Subjects.—Seventy-four students participated in the experiment as *Ss* to fulfill one of the course requirements in introductory psychology. There were 16 *Ss* in each of four experimental groups and 10 *Ss* in the control group.

Procedure.—The general procedure, with but slight modifications, was that employed by Rhine and Silun (1958) and also described by Rhine (1958). Briefly, it was as follows: Subjects were administered the task individually by use of a set of 50 4 × 6 in. cards on each of which was a single word, in primary type, representing a "trait" or "characteristic." The instructions were that *E* wanted to find out whether *Ss* could make predictions about the characteristics of each of two little known cultures from the kinds of information typed on the cards. Since *E* wanted *S* to start out with as little information as possible the actual names would not be used, but instead the two words CEB and LOJ would be used to label the cultures. The "names" were picked arbitrarily from Glaze (1928) association values of 27%. The *S* was told to look at one trait at a time and predict to which culture it belonged by first saying the name of the culture and then the trait (e.g., *S* might say, "CEB-GRANDIOSE"). Following each such prediction *E* announced the correct combination as a basis for *S* to evaluate his performance. If *S* was correct he went on to the next card. If incorrect, *S* repeated the correct combination and then proceeded with the next prediction. The *S* went through 50 trait names in order and then repeated the series a second time. A record of *S*'s errors was kept by *E*.

Following this procedure *S* rated each of the newly developed concepts, CEB and LOJ, separately on 20 semantic differential scales in exactly the same form and order used by Jenkins, Russell, and Suci (1958, p. 690). These scales had been selected by Jenkins et al. (1958) because the factorial loadings were highest on the dimension sampled by each scale and low on all other factors.

The treatments differed with respect to the concept-formation tasks required of the experimental groups. The general format of the task, however, was exactly the same for all groups. It was comprised of 50 different trait names from which *S* was to make his prediction; 25 of these traits were to be associated with one of the nonsense syllables and were classifiable on either the evaluation or potency dimension depending upon the experimental variation in which the list was to be used; the remaining 25 traits were represented by neutral words and were not classifiable on the respective dimensions of meaning. The specific words and the serial position of each neutral word were exactly the same for each of the experimental groups and the control group. The position occupied by the traits classifiable on the evaluative or potency dimension was the same in all groups. For the control group, another set of neutral words was used in place of the experimental variations of the word lists.

The nonsense syllable used to designate the critical concept was associated in Group C_{NE} with unfavorable (negative-evaluative) words; in Group C_{PE} with favorable (positive-evaluative) words; in Group C_{HP} with words high in value on the potency dimension; in Group C_{LP} with words low in value on the potency dimension; and, in the control group (Group C_N) with words having neutral meanings. In all groups, as indicated above, the other nonsense syllable was associated with words having neutral meanings, the same words being used for all groups. The data related to the learning of the neutral words will be designated by the symbols N_{NE}, N_{PE}, N_{HP}, N_{LP}, and N_N for the respective experimental and control groups.

In the positive evaluative list were included such words as GENUINE, RELIABLE, and CHARITABLE and in the negative evaluative list were such words as DESTRUCTIVE, DIRTY, and CHEAT. The words in both of these lists were selected from those used by Rhine[2] and Silun who chose only those words for which there was 90% or more agreement among *Ss* on the classification of the words as "very desirable" or "very undesirable" traits.

[2] R. J. Rhine, personal communication.

The words used in the high and low potency meaning lists and in the neutral list were selected on the basis of mean concept values on the masculine-feminine and hard-soft scales published by Jenkins et al. (1958). Words for the high potency list were selected if the mean values approximated 3.00 or less. The average scale values of all words finally used were 2.24 and 1.96, respectively, on the two scales. The opposite ends of the same scales were used in the selection of words with low potency meaning. The average scale values for the 25 low potency words on the two scales were 5.19 and 5.36, respectively. Illustrations of words having high potency meanings are RAPID, BASE, ANGER, and FIRE; of words having low potency meanings are SMOOTH, LEISURELY, and CALM.

Examples of words in the neutral list are FAR, ARCHED, LINGERING, and CONTINUOUS. The words in this category were chosen from the range of *Atlas* values between approximately 3.00 and 5.00 on each of the three dimensions concerned, and insofar as possible on the remaining 14 scales. In practice, there were few departures from this general rule. The average scale values of the words so selected were 3.87 on the evaluative scales, 4.05 on the potency scales, and 3.86 on the activity scales.

The concepts selected for each list thus were chosen from a specified portion of the rating scale for the manipulated dimensions. However, as noted by the authors (Jenkins et al., 1958), and as is evident from the examples above, the *Atlas* does not permit the selection of words representing factorially pure meanings.

The analyses to be reported below are based on *Ss*' ratings on 6 of the 20 scales used. The cruel-kind and good-bad scales were used in measuring the evaluative dimension, and the hard-soft and masculine-feminine scales for measuring the potency dimension. These particular scales were selected because they represent the most highly loaded dimensions involved in learning the critical concepts by the respective experimental groups. The active-passive and excitable-calm scales were used, in addition, to measure the oriented-activity dimension. Inasmuch as no attempt had been made to vary the activity dimension it was assumed that comparisons of data from the evaluative and potency scales with that gathered from this measure might provide a supplementary control. In practice, however, it was virtually impossible to select words from the available lists which were high on either the potency or evaluation dimension and low on the activity dimension. In all comparisons, analyses, and combinations of scales within a dimension, the favorable pole of the scales was aligned with a rating of 7 and the unfavorable pole with a rating of 1.

The nonsense syllable associated with the critical concept and that with the neutral concept were counterbalanced within each group. In none of the groups were differences in learning, and in the consequent ratings for concepts learned to the syllables CEB and LOJ, significant ($P > .05$) by t test. Since the nonsense syllables were found not to have a significant effect on either learning or ratings the two subgroups were combined

TABLE 1

MEAN RATINGS COMPARED WITH MEAN ATLAS VALUES OF WORDS USED IN CONDITIONING THREE CONCEPT-ATTITUDES

Semantic Differential Scale	Concept-Attitude					
	Neutral (NS)		High-Potency (CHP)		Low-Potency (CLP)	
	Mean Rating	Mean *Atlas*	Mean Rating	Mean *Atlas*	Mean Rating	Mean *Atlas*
Cruel-Kind	4.4	4.2	2.3	2.9	6.1	5.9
Good-Bad	3.4	3.5	4.2	4.0	2.1	2.3
Masculine-Feminine	3.9	4.0	1.7	2.2	5.9	5.2
Hard-Soft	4.6	4.1	1.9	2.0	5.5	5.4
Active-Passive	2.2	3.9	1.7	2.2	3.5	3.9
Calm-Excitable	4.0	3.8	5.5	5.0	1.9	2.8
D	1.80		1.08		1.24	

Note.—In this and the following tables, N in Group NS is 10, and N in each other group is 16.

TABLE 2

MEANS AND *SD*s OF CONCEPT RATINGS ON EVALUATIVE, POTENCY, AND ACTIVITY SCALES

Concept-Attitude	Rating Scale					
	Evaluative		Potency		Activity	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Neg. Evaluative (C_{NE})	1.44	0.58	1.72	0.72	2.07	1.12
Neutral (N_{PE})	3.34	1.22	3.15	1.37	3.69	1.63
Neutral (N_N)	4.80	1.55	4.25	1.46	3.05	1.21
Neutral (N_{NE})	5.88	0.80	5.19	1.02	5.59	1.27
Pos. Evaluative (C_{PE})	6.56	0.71	4.15	1.39	4.09	1.24
High Potency (C_{HP})	3.03	1.56	1.76	1.20	1.99	1.30
Neutral (N_{LP})	3.69	1.05	2.69	1.23	3.31	1.58
Neutral (N_N)	4.80	1.55	4.25	1.46	3.05	1.21
Neutral (N_{HP})	4.86	1.52	5.26	1.10	4.99	1.63
Low Potency (C_{LP})	6.00	0.81	5.72	1.14	4.79	1.25

within each experimental group and in the control group for all subsequent analyses.

RESULTS

In Table 1 *Ss*' ratings, on six semantic differential scales, of C_{HP}, C_{LP}, and N_N are compared with the mean of the *Atlas* values for all 25 words that were used in the conditioning of each attitude in these groups. The difference, *d*, between the mean ratings of the conditioned concept-attitudes and the mean *Atlas* values was found for each of the scales. These were combined into a single measure, *D*, a descriptive statistic, of the distance between the conditioned concept-attitude and the *Atlas* scale values using the formula $D = \sqrt{\Sigma d^2}$. While the significance of *D* cannot be tested directly it should be noted that the maximum value of *D* with the six scales is 14.7. Thus, the small values reflect essential agreement between *Ss*' ratings and those published in the *Atlas*. Inasmuch as the negative-evaluative and the positive-evaluative lists were obtained from another source, comparable comparisons were not possible.

In Table 2 are presented the means and *SD*s of ratings given the critical and neutral concepts on each of the three dimensions, evaluative, potency, and activity, by each group. The ratings on the two individual scales, on which each dimension was measured, were combined for this and all subsequent analyses to obtain a total "factor" score. Similarly, since the difference in ratings between C_N and N_N was not significant ($P > .05$, sign test) the two were combined and are represented as N_N.

While there is considerable generalization among factors to be noted in the data summarized in Table 2, the most consistent and strongest effect occurs in the factor or dimension on which the word lists had been selected. Also apparent is the effect on the ratings of the neutral concept by the principle or dimension involved in the conditioning of the critical concept. The scores are arranged to highlight this effect, the arrangements being identical for the two dimensions.

In Table 3 are presented the distances for each critical concept compared with the relevant neutral con-

TABLE 3

DISTANCES (*D*) BETWEEN RATINGS MADE OF THE CRITICAL AND NEUTRAL CONCEPTS

Critical Concept	Distance (of Critical Concept) from Neutral Concept Learned Concomitantly With:		
	Critical Concept	Neutral Concept	Opposing Critical Concept
C_{HP}	7.44	4.36	2.37
C_{LP}	6.12	3.97	2.99
C_{PE}	5.24	3.15	3.07
C_{NE}	8.79	6.27	4.01

cepts (either that learned with the critical concept or with the opposing critical concept). In every instance, the lowest values of *D* are between the critical concepts and the neutral concepts conditioned with the opposing critical concept. The largest values of *D*, on the other hand, are between the critical concepts and the respective neutral concepts conditioned with them. The data in this table reflect the specific effects of the meaning dimension on the ratings of the concept. In every instance where the difference was not significant ($P > .05$) it was either on the activity scale or on the scale for the unrelated factor.[3]

The comparisons among the neutral concepts learned in conjunction with the various critical concepts are presented in Table 4. It is clear, from this as in the previous table, that the neutral concept is strongly affected by the concurrent learning of another concept with a strong meaning dimension. The largest *D*s are to be found between the neutral concepts learned under opposing conditions.

All *D*s shown in Tables 3 and 4 were tested indirectly on each dimension by use of nonparametric statistics.[3] The sign test was used for all related samples (e.g., in testing the significance of the difference between ratings of C_{NE} and N_{NE}). The Fisher exact probability test was used for testing the difference between ratings of all independent samples (e.g., in testing the difference between the ratings of the C_{NE} and N_{PE}). The median for all comparisons was between the .01 and the .001 level of significance. These differences indicate: (*a*) that the critical concepts were developed in accordance with the meaning dimension of the words that were associated with them; (*b*) that, on the dimension being conditioned, the neutral concepts were substantially different from each other; and (*c*) the ratings made of the neutral concepts were most like the ratings of the opposite critical concept.

A further analysis demonstrating the dependence of the rating of the neutral concept on the rating of the respective critical concept was accomplished by correlating the two ratings of all *S*s ($N = 74$) in all groups. The Pearson *r*'s, all high and significantly greater than zero ($P < .01$), are as follows: $r = -.73$

TABLE 4

DISTANCES (*D*) BETWEEN RATINGS MADE OF THE NEUTRAL CONCEPT LEARNED UNDER THE SEVERAL EXPERIMENTAL CONDITIONS

Comparison	*D*	Comparison	*D*
N_N vs. N_{HP}	3.75	N_N vs. N_{NE}	4.29
N_N vs. N_{LP}	2.62	N_{LP} vs. N_{HP}	5.17
N_N vs. N_{PE}	3.44	N_{NE} vs. N_{PE}	4.78

[3] Two tables presenting the levels of significance on each dimension for the comparisons shown in Tables 3 and 4 have been deposited with the American Documentation Institute. Order Document No. 6863 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

for evaluative ratings; $r = -.79$ for potency ratings; and $r = -.76$ for activity ratings.

Differences between experimental groups in the total number of errors made during the learning phase were tested for significance by analysis of variance for simple randomized groups. The mean number of errors for each group and the F for the comparison among groups are presented in Table 5. The comparisons among means were made by Duncan's multiple range test. These comparisons revealed no significant differences in learning errors between Groups C_{NE} and C_{PE}, or between Groups C_{HP} and C_{LP}, on the critical concepts. Group C_{PE} did learn with significantly ($.05 > P > .01$) fewer errors than either Group C_{HP} or Group C_{LP}, however. In the comparisons involving the learning of the neutral concepts, significantly ($P < .05$) fewer errors were made in learning N_{PE} than in learning N_{HP}, N_{LP}, or N_{NE}. No other differences were found to be significant. Since these differences were not related in any systematic way to the concepts being compared, it is doubtful that the degree of learning affected the ratings made. On the other hand, in every instance, the critical concept was learned with fewer errors than was the neutral concept. Over all groups, approximately 40% more errors were made in learning the neutral concepts, running as high as twice as many errors in the groups learning words varied on the evaluative meaning dimension. In the control group ($N = 10$) an average of 23.6 ($SD = 4.63$) errors were made in learning one of the neutral concepts and an average of 23.7 ($SD = 3.82$) errors in learning the other neutral concept. This result is not significantly different from expectations based on chance.

TABLE 5

MEAN ERRORS IN LEARNING THE CRITICAL AND NEUTRAL CONCEPTS IN EACH EXPERIMENTAL GROUP

Group	Critical Concept		Neutral Concept	
	Mean	*SD*	Mean	*SD*
C_{NE}	8.50	7.39	15.81	6.03
C_{PE}	5.52	4.97	11.44	4.47
C_{HP}	13.44	4.58	17.19	3.42
C_{LP}	11.13	6.17	18.81	6.52
$F(df = 3/60)$	4.73		5.34	
P	<.01		<.01	

DISCUSSION

The evidence supports the general hypothesis that attitudes can be developed by associating symbols (nonsense syllables) with words having in common an explicit meaning dimension. Here the assumption is that through the past history of S positive-meaning words were associated with some mediator like that representing good and negative-meaning words with some other mediator such as bad. Similarly, high potency words are, by a similar process, associated with some common mediator like hard and low potency words with a mediator such as soft. In the present experiment the evaluative dimension of the stimulus-traits was most distinct for the groups learning the negative or positive characteristics and the potency dimension was most distinctive for the remaining two groups. During the learning process it is assumed that the abstraction by S of the common element mediated the rating of the critical concepts on the semantic differential scales. Our results thus support those of previous investigators, particularly the conclusions of Rhine and Silun (1958). At this point in the discussion, the major difference between the present study and the earlier investigations is that in the present study the procedure was extended to dimensions of meaning other than the evaluative dimension. Thus, it was demonstrated that the potency

dimension of meaning identified by Osgood et al. (1957) does provide a basis for semantic mediation as does the evaluative dimension. In addition, the semantic differential scales were used in rating the concepts, a procedure closer to that traditionally used in the measurement of attitude than the continued elicitation of a successive series of traits used by Rhine and Silun (1958).

The extreme stability of the ratings of the various concepts developed can almost be described as dramatic. As indicated, the present ratings were nearly identical with those arrived at through an average of the *Atlas* values of the words used in conditioning both potency and neutral meaning to the nonsense syllables. Furthermore, in a very real sense, our four experimental groups comprise four replications. In each case, there was no question that the ratings of the critical concepts were in the direction and of the approximate magnitude that would be predicted on the basis of the words used.

The major hypothesis of the present study was that the ratings made as a consequence of the conditioning of the critical concept would be inversely related to the ratings made of the syllable associated with the neutral words. This hypothesis was supported completely in each of the experimental groups. Not only is such a contrast effect present, but it too appears to be an extremely consistent and powerful effect. The data compare favorably with those obtained through the use of simple stimuli (Sherif et al., 1958) and through the studies of contrast occurring as a result of the comparison of source-content (Hovland et al., 1957) variables in attitude formation.

The findings clearly indicate that the evaluation of the neutral concept is dependent upon the meaning abstracted in the conditioning of the critical concept. If degree of learning is used as a measure of strength of concept-attitude development (Rhine and Silun, 1958), it is apparent that the stronger of the two concepts (in this instance, the critical concept) is the anchor. The high correlation between the ratings of the critical concept and its corresponding neutral concept indicates that the greater changes occur at the extremes and the lesser changes (assimilation?) occur at the intermediate rating levels. Nevertheless, while the over-all contrast effects are strong, in no case do they equal in intensity the ratings of the critical concept-attitudes.

In the interpretation of the contrast effects a direct-opposite associative effect is highly favored in accounting for the evaluation of the neutral concepts subsequent to the training trials. This is consistent with Osgood's (1953, pp. 708–711) statement that opposites cannot be based upon semantic mediation processes, but are the result of highly overlearned verbal habits. Nevertheless, this does not preclude the possibility of a strong self-instruction habit, as a product of learning-how-to-learn, existing as a mediating process (see Campbell, Lewis, & Hunt, 1958, for example). Thus *Ss* in a partially ambiguous situation behave as though a verbal rule such as that of "select-the-opposite" is particularly adaptive in an unstructured situation requiring evaluative responses. Once such a verbalization permits the structuring of a situation there follows the elicitation of the specific combination of opposites having the strongest habit strength within a hierarchy of possible responses. The particular hierarchy employed appears to be dependent upon a contextual definition based on the anchor point (Di Vesta & Bossart, 1958; Phillips, 1958).

There is one further implication of the present findings that concerns the design of investigations in this area. In the majority, if not all, studies on the conditioning of attitudes (Staats & Staats, 1957, 1958; Staats, Staats, & Biggs, 1958) and on the development of concept-attitudes (Rhine & Silun, 1958) the possibility of the contrast effect appears to have been neglected. Important in this respect was the fact that *Ss* learned two concepts, concomitantly, one to positive evaluative words and the other to negative evaluative words, for ex-

ample. In view of the present evidence, the results of said procedure would be to accentuate the stronger of the two evaluative poles but the rating of the weaker of the two concepts might well be the result of a direct (i.e., contrast effect) rather than mediated association. Under the circumstances it would appear that a study of two concepts that are learned concomitantly may be an inadequate procedure. Because of the possible interaction between the two, separate treatments in each group, rather than a covariance design, would be desirable.

On the basis of the Rhine and Silun (1958) study it would be expected that attitude development and attitude strength depend in large part on consistency of reinforcement (degree of learning). Provision in the present study was made for complete learning of the critical concepts by determining through pilot studies, the number of trials required for an abstraction to be made by *Ss*. The analyses indicate that while differences in the mean number of errors between groups were to be found in the learning of the critical concepts and of the neutral concepts, the errors did not occur in a way that was related to the analyses performed. Thus, the differences in errors between those groups for which the ratings were directly compared (e.g., negative-evaluative vs. positive-evaluative) were in only one case (between the errors made in learning the neutral concepts learned respectively with the positive evaluative and negative evaluative critical concepts) significant. On the other hand, in every case the critical concepts were learned with fewer errors than the respective neutral concepts. Thus, some support is given to the hypothesis that essential contrast may be a superficial definition of the concept depending upon a direct, rather than mediated, association. Nevertheless, the phenomenon appears as a strong and potentially important influence in the learning of attitudes. The conditions of its demonstration in the present study calls attention to a heretofore neglected source of the occurrence of the "boomerang" effect in the development of attitudes.

Summary

Each of four groups ($N = 16$ each) of *Ss* learned one concept-attitude based on words strongly representing the evaluative or potency dimension of meaning and another concept-attitude based on words having neutral meanings on the respective dimension being manipulated. A control group ($N = 10$) learned two concept-attitudes, both with neutral meanings.

Concept-attitudes are shown to be developed specific to the meaning dimension involved in the words used in conditioning. Furthermore, in support of the major hypothesis, concept-attitudes learned to neutral meanings, although conditioned to exactly the same words in all groups, were found to be rated by *Ss* in direct opposition to the respective critical concept.

A mediation hypothesis was suggested as the best explanation of the changes in the critical concept. The contrast effect was explained in terms of the critical concept providing an anchor to which a direct-opposite response was applied for the rating of the less structured neutral concept. It was indicated that in previous studies of the verbal conditioning of attitudes some of the effect attributed to the mediation hypothesis may be little more than the formation of essential contrast. Note was also made that the contrast effect determined through contextual definition (while holding communicator effects constant) provides still another explanation of the so-called "boomerang" effect in the development of attitudes.

References

Campbell, D. T., Lewis, N. A., & Hunt, W. A. Context effects with judgmental language that is absolute, extensive, and extra-experimentally anchored. *J. exp. Psychol.*, 1958, **55**, 220–228.

Di Vesta, F. J., & Bossart, P. The effect of sets induced by labelling on the modification of attitudes. *J. Pers.*, 1958, **26**, 378–387.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

Harvey, O. J., & Caldwell, D. F. Assimilation and contrast phenomena in response to environmental variation. *J. Pers.*, 1959, **27**, 125–135.

HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55,** 244–252.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, **71,** 688–699.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PHILLIPS, L. W. Mediated verbal similarity as a determinant of the generalization of a conditioned GSR. *J. exp. Psychol.*, 1958, **55,** 56–62.

RHINE, R. J. A concept-formation approach to attitude acquisition. *Psychol. Rev.*, 1958, **65,** 362–370.

RHINE, R. J., & SILUN, B. A. Acquisition and change of a concept attitude as a function of consistency of reinforcement. *J. exp. Psychol.*, 1958, **55,** 524–529.

SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55,** 150–155.

STAATS, A. W., & STAATS, C. K. Attitudes established by classical conditioning. *J. abnorm. soc. Psychol.*, 1958, **57,** 37–40.

STAATS, A. W., STAATS, C. K., & BIGGS, D. A. Meaning of verbal stimuli changed by conditioning. *Amer. J. Psychol.*, 1958, **71,** 429–431.

STAATS, C. K., & STAATS, A. W. Meaning established by classical conditioning. *J. exp. Psychol.*, 1957, **54,** 74–80.

(Received September 12, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 545-551

THE SPACING OF SEQUENTIALLY DEPENDENT TRIALS IN PROBABILITY LEARNING [1]

JEAN P. CHAPMAN [2]

Northwestern University

The present study is an investigation of some implications of the statistical model developed by Estes (1955) and Burke and Estes (1957) and a recent extension of this model by Witte (1959). More specifically it is a study of the effect of massed and distributed trials on response probabilities when the reinforcing events are sequentially dependent. Markov series of two-choice events are used in a Humphrey's-type verbal conditioning situation.

Recently there have been several studies using a sequentially dependent reinforcement schedule. Hake and Hyman (1953), investigating a different problem, studied both independent and dependent series of experimental events. Engler (1958) tested some predictions from the Bush-Mosteller model, using both independent and dependent series. She found that asymptotic marginal response probabilities approximately match marginal stimulus probabilities, but that asymptotic conditional response probabilities appear to exceed conditional stimulus probabilities except when the latter are extremely low.

A recent investigation by Witte (1959) used an extension of the Burke and Estes theory. To this theory Witte added a stimulus trace hypothesis to predict asymptotic behavior when trial spacing is distributed and the reinforcements are sequentially dependent. Witte predicted that conditional response probabilities should match the sequential dependencies between rewards only when trial spacing is massed. When the trial spacing is sufficiently distributed, these probabilities should approach the marginal probability for most marginal and conditional probability specifications. Witte tested this using rats in a T maze and found that the terminal conditional response probabilities for *S*s at each intertrial interval differed significantly in the direction predicted by the theory. However, the theory did not adequately describe the response probability at all points on the trial spacing continuum.

Witte presented a formula (his Equation 4) for the probability $p_{11}(n+1,t)$ of an A_1 response after trial n, given an E_1 occurrence on trial n and an intertrial interval of t. The solution of his Equation 4 is:

$$p_{11}(n+1,t) = \frac{1}{N_e + N_1(t)}\{N_e[(1-\theta)^n(F_e(1) - \bar{\pi}) + \theta(1-\bar{\pi})] + N_1(t)[(1-\theta\pi_{11})(1-\theta\bar{\pi})^{n-1}(F_1(1) - \pi_{11}) + \theta\pi_{11}(1-\pi_{11}) + \pi_{11}]\}, \quad [1]$$

with

$$Pr(E_1) = \bar{\pi} = \frac{\pi_{21}}{1 - \pi_{11} + \pi_{21}} \quad [2]$$

[1] This paper is based on a PhD dissertation at Northwestern University. The author is indebted to John Cotton for invaluable assistance and advice in directing the dissertation, and to William K. Estes for his suggestions on the theoretical portion of this study.

[2] Now at University of Kentucky.

Thus the conditional response probabilities tend asymptotically to the corresponding conditional stimulus probabilities plus or minus "correction terms" of magnitude smaller than θ, and marginal response probabilities tend asymptotically to marginal stimulus probabilities plus or minus "correction terms" of magnitude smaller than θ.

Inspection of this formula reveals that it is a kind of weighted average of the Burke and Estes model for massed and distributed trials. Comparison of the Burke and Estes formulation with that of Witte reveals that the two models make a different prediction for massed trials except in the case of N_c/N_1 equal to zero. In order for Witte's model to be applicable to both massed and distributed trials this ratio must be greater than zero.

In the present investigation a comparison of these two models was made for data obtained with massed trials. A recent study by Anderson (1960), published since the completion of this work, tested the Burke and Estes (1957) model using data from a two-choice probabilistic task with sequentially dependent stimulus events. He found that the Burke and Estes model predicted terminal response levels quite well for the larger values of π_{11} but did poorly for the very low values.

Inspection of the model by Witte shows an important characteristic of this model. Each trial is affected only by what has occurred on the previous trial and by the length of the intertrial interval. This leads to the prediction that if, after a long period of massed trials spaced trials were given, the massed trials would have no effect on the spaced trials after the first spaced trial, and the response probabilities would shift immediately to the terminal values for the spaced trials. For longer spacing intervals the terminal response probabilities should be (for most marginal and conditional probability specifications) closer to the marginal probabilities plus or minus the correction terms.

As a test of this prediction *S*s were given a long series of massed trials with the reinforcements sequentially dependent. After the massed trials, each *S* was given a series of distributed trials with the same probability specifications as the massed trials.

Method

Apparatus.—A black board 39 in. × 31 in. separated *S*'s apparatus from the apparatus visible to *E*. On *S*'s side of the screen were two brown boxes 4 in. × 4 in. × 2 in. mounted 4 in. apart. On the vertical front face of each box was mounted a red light 1 in. in diameter and below it a toggle switch. From these boxes wires ran under the screen to the apparatus on *E*'s side of the screen. On this side of the screen were mounted a toggle switch which started a buzzer, two 15-w. bulbs each of which was lighted by the toggle switch on the appropriate box on *S*'s side of the screen, and two toggle switches each of which lighted the red light on the appropriate box on *S*'s side of the screen. The toggle switch on each of *S*'s boxes turned off the buzzer as well as turned on the appropriate light on *E*'s side of the screen. The two boxes in front of *S* and the corresponding lights were arbitrarily labeled E_1 and E_2. The *S* was not aware of this labeling. The E_1 box occurred on the right for half the *S*s and on the left for the other half.

Subjects.—The *S*s were 120 students from courses at the University of Chicago and the University of Kentucky. They were assigned to the 12 groups using a sequence from a random numbers table, with the restriction that 10 *S*s occur in each group. None of the *S*s had previously served in an experiment of this type.

Procedure.—Each *S* was given 500 trials. The first 480 trials for all *S*s were massed (one every 5 sec.) and were run without a break. There were four sets of marginal and conditional probability specifications used as shown in Table 1. Thirty *S*s were contained in each of the four groups corresponding to the four probability specifications. After the 480 massed trials, each of the four groups was divided into three subgroups. These three subgroups were given the 20 additional trials with intertrial intervals of 8 sec., 1 min., and 5 min., respectively. These intertrial intervals were from the presentation of the reinforcement until the

sound of the buzzer starting a new trial, and were timed with a stop watch. The probability specification was the same for a given S throughout all 500 trials.

The E signaled the beginning of each trial by sounding a buzzer. The S then made his choice by pressing the toggle switch on the appropriate box. This stopped the buzzer and lighted the bulb which indicated his choice to E. Then E indicated the correct choice to S by pressing the toggle switch which lighted the light on the appropriate box in front of S.

The orders of presentation of the reinforcements were constructed to comply with four sets of marginal and conditional probability specifications. These probability specifications were chosen to enable comparison of the obtained results with those of Engler (1958). She had four groups that had the same conditional probability specifications as did the present study (i.e., $\pi_{11} = .70$ and .25) but had different marginal stimulus probabilities ($\bar{\pi} = .25$ and .50). A different sequence was made up for each S using tables of random numbers (Rand Corporation, 1955; Snedecor, 1946; Walker & Lev, 1953) each sequence being constructed according to the probability specifications for the group to which the S was assigned.

The theory specifies that these sequences be truly random. In order to comply with this specification the sequences were constructed in the following manner. The experimental event occurring on Trial 1 was determined by using a table of random numbers taking into account the value of π_1. The experimental event occurring on Trial 2 was then determined by using a table of random numbers taking into account the value of π_{i1}, $i = 1, 2$ where E_1 occurred on Trial 1, and so forth for all n trials. The resulting sequences do not agree exactly with the specified probabilities, but the group of sequences represents a random sample from the population of such sequences.

Results

The learning parameter θ was estimated from the marginal data using the following equation:

$$\theta = \frac{\pi_1 - p_1(1)}{v(\pi_1) - \sum_{n=1}^{v} p_1(n+1)} \quad [3]$$

where v is the number of trials used to estimate θ. Theta was estimated for each group, and the obtained values were .037, .062, .028, and .024 for the four groups, respectively.

TABLE 1

Marginal and Conditional Probability Specifications for Reinforcement Sequences and Observed Asymptotic Marginal Response Probabilities

Group	$Pr(E_1)$	$Pr(E_1\|E_1)$	$Pr(E_2)$	$Pr(E_2\|E_2)$	Observed Asymptote
I	.20	.70	.80	.9250	.18
II	.20	.25	.80	.8125	.13
III	.35	.70	.65	.8384	.33
IV	.35	.25	.65	.5962	.32

Massed data.—The obtained response probability curves are shown in Fig. 1. The proportions of A_1 responses following an E_1 occurrence are plotted for blocks of 20 trials. These curves are negatively accelerated and nearly asymptotic in the later trials, and they are quite variable, particularly in the later stages of practice.

The sections of Fig. 1 also contain the predicted response curves from the two theoretical models, that of Burke and Estes (1957) and that of Witte (1959). Comparison of the obtained curves with those predicted by Burke and Estes shows that the performance for Groups I and III is almost always well above the predicted curve. The terminal response values for these two groups also are above the predicted asymptotes. These groups both had reinforcement sequences constructed with a conditional probability (π_{11}) of .70.

Groups II and IV had reinforcement sequences with the conditional probability of an E_1 of .25, given an E_1 on the previous trial. The Burke and Estes (1957) predicted curve for Group II fits the data fairly well. This is true particularly in the early stages of practice, but there is some indication that the response proba-

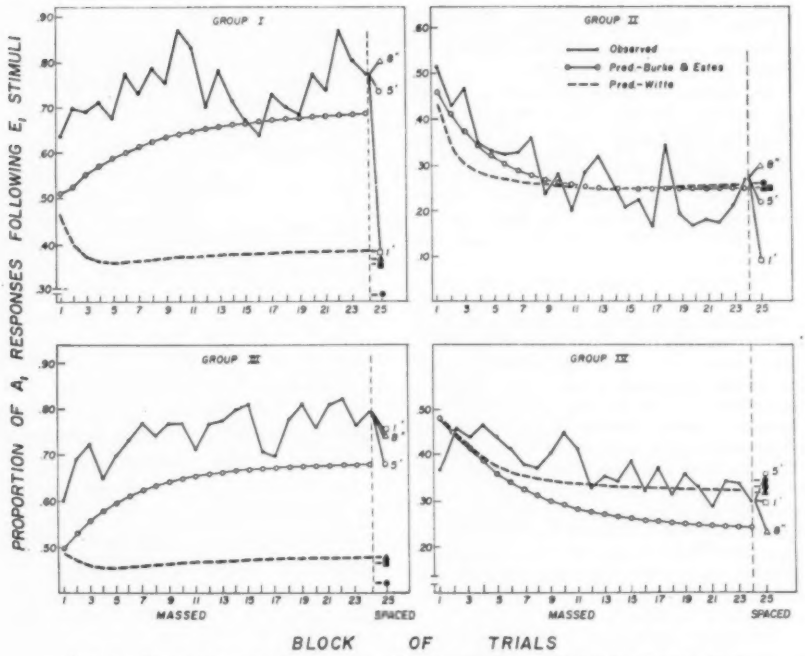


FIG. 1. Predicted and obtained conditional response probabilities for the four groups on massed and distributed trials.

bilities are becoming asymptotic beyond the predicted value. On the other hand the response curve for Group IV appears to be stabilizing at a value less than the predicted value.

A comparison of the Witte (1959) predictions with the obtained curves shows that Group IV produced the data that best matched the predicted curve. This group was the one used to estimate the necessary constants, N_c/N_1 and j_1.[3] In fact it was the only group which made possible such estimation. (As described earlier this model is a kind of weighted average of the Burke and Estes [1957] formulation for massed and completely distributed trials. In order for the Witte model to be used, it was necessary for the obtained terminal response data to be either equal to the predicted asymptote for the Burke and Estes massed trial model, or between the asymptotes for the massed and distributed models. This condition was met only in Group IV.) The Witte prediction seems to do about as well as the Burke and Estes prediction for Group II, particularly in the later stages of practice.

Figure 1 shows the N_c/N_1 ratio estimated from Group IV to be useless in predicting the performance of Groups I and III. Making a new estimate of N_c/N_1 to make this model fit the obtained data from these groups results in a negative ratio. Theoretically this would say that there are more elements (N_1) avail-

[3] The method for estimating the constants may be found in the author's dissertation which has the same title as the present paper. The values were estimated using the asymptotic response probabilities. The obtained values were $N_c/N_1 = 1.91$ and $j_1 = .0034$.

able after an E_1 occurrence than are present in the entire set of N_1+N_e elements, or that there are a negative number of elements available in the S_e set. These outcomes do not have much intuitive appeal as the theory now stands, and the incorporation of them would mean a major revision of the theory.

The left half of Table 2 shows S's terminal response probabilities based on the last 100 trials and the predicted asymptotes from both models. The difference between the obtained asymptotes and the Burke and Estes (1957) predicted asymptotes was tested using a t test for all groups. The differences for Groups III and IV were significant ($P < .01$) but the differences for Groups I and II were not significant ($P = .10$ and $P = .60$).

Distributed trial data.—The obtained data for the 20 distributed trials for the four groups is shown at the far right of each section of Fig. 1. The Witte theory predicts that after a series of massed trials the proportion of A_1 responses following E_1 stimuli should immediately switch to a point closer to the marginal probability specification, plus or minus a correction term. The prediction is somewhat different for Group II where the marginal and conditional probability specifications are close together. The predicted massed asymptote for this group is .238 and the completely distributed asymptote is .249, with the conditional probability equal to .25 and the marginal probability equal to .20. This switch in the effect of spacing is due to the magnitude of the correction term for this group.

The Witte predictions for the three intertrial intervals for each group are shown at the far right of each section of Fig. 1. The groups with the greatest predicted switch are I and III, the groups with the greatest discrepancy between π_{11} and $\bar{\pi}$. These groups show some tendency to switch to the predicted values but not immediately as is required by the theory. The deviations of these groups from the predicted values were tested using the Wilcoxon (1949) signed ranks test. The results of this test are shown in Table 2. The 8-sec. and 5-min. values are both significantly different from the predicted values for Group I. For Group III both the 1-min. and 5-min. intervals produced significant differences. Groups II and IV are a little harder to interpret since the change predicted by the theory is not very great.

TABLE 2

OBSERVED AND PREDICTED ASYMPTOTIC PROBABILITIES OF A_1 RESPONSES FOLLOWING E_1 STIMULI

Group	Massed Trials			Distributed Trials		
	Obs.	B&E Pred.	Witte Pred.	Interval	Obs.	Pred.
I	.79	.69	.39	Massed		.69
				8 sec.	.80*	.38
				1 min.	.38	.37
				5 min.	.74*	.30
				Completely spaced		.23
II	.20	.24	.25	Massed		.24
				8 sec.	.30	.25
				1 min.	.10*	.25
				5 min.	.22	.25
				Completely spaced		.25
III	.79	.68*	.48	Massed		.68
				8 sec.	.75	.47
				1 min.	.76*	.46
				5 min.	.68*	.42
				Completely spaced		.37
IV	.32	.25*	.33	Massed		.25
				8 sec.	.24	.33
				1 min.	.26	.33
				5 min.	.35	.35
				Completely spaced		.37

Note.—Massed data were tested against the Burke and Estes prediction.

* These observed values differ from predicted with $P < .05$.

TABLE 3

COMPARISON OF OBTAINED MARGINAL RESPONSE PROBABILITIES AT 240 TRIALS OF ENGLER'S DATA AND PRESENT DATA

Conditional Probability	Marginal Probability			
	.20 (Present)	.25 (Engler)	.35 (Present)	.50 (Engler)
.25	.179	.33	.324	.493
.70	.157	.22	.322	.498

In fact the predicted distributed trial values are close to those predicted for the massed trials. Inspection of the results for these two groups reveals that the obtained values are scattered about the predicted values. Only one of the six differences tested for these two groups is significant (Group II: 1 min.).

For each of the four groups the Kruskal-Wallis H test was used to test whether the three subgroups having different intertrial intervals differed from each other in response probability. The differences were not significant for any of the four Groups. Group I showed a trend toward significance ($H = 5.59$, $P \cong .07$) with the 1-min. group being the deviant one.

Each of the spaced groups was compared with its respective massed asymptote to test whether there was a significant change in conditional response probability when trial spacing was introduced. These differences were tested using a sign test. Only one of the 12 subgroups showed such an effect. The 1-min. condition for Group I showed a significant change from the massed to the spaced condition.

Marginal data.—The theory of Burke and Estes (1957) predicts the marginal response probability equal to $Pr(E_1)$. Table 1 shows the obtained and predicted asymptotic marginal response probabilities. The obtained asymptotes are based on the last 100 trials. This table shows the obtained marginal responses to be consistently below those predicted by the theory. In order to compare the present data more directly with these of Engler (1958) the response probability values at 240 trials were computed from the present data. These values at 240 trials are compared with Engler's findings in Table 3. At this stage of practice the present data match the marginal probability specifications about as well as does the Engler data.

As an indication of the effect of conditional probability specification of the series on marginal response behavior, a 2 × 2 analysis of variance of marginal response probabilities was computed. No difference was found in marginal responses due to conditional probability specification ($F = .205$, $df = 1/116$). The interaction of marginal and conditional probability specifications also was not significant ($F = 1.94$).

DISCUSSION

The results of this study show that with sequentially dependent massed trials, the asymptotic conditional response probability does not uniformly match the conditional stimulus probability. Consistent with the results of Engler (1958), in some series when the conditional probability was high ($\pi_{11} = .70$) and the trials massed, the conditional response probability of an A_1 response following an E_1 stimulus exceeded the series specification. In the groups where the series specification was low ($\pi_{11} = .25$) the groups approximately matched the predicted asymptotes.

This study was designed also to test some inferences from two models, one of Burke and Estes (1957) and a modification of this model by Witte (1959). The results confirmed the finding by Anderson (1960) that the mathematical predictions showed some inadequacies in predicting the performance of human

Ss in a verbal conditioning situation with sequentially dependent reinforcements.

The mathematical predictions about the distributed trials were not upheld in the two groups for which a large shift was predicted. The Witte model predicted that if, after a series of massed trials, spaced trials were given, the response probabilities would immediately shift to the terminal values for the spaced trials, the amount of shift determined in part by the size of the difference between the marginal and conditional stimulus probabilities. This immediate shift was not found in the data, though there seemed to be some indication of a trend of *Ss*' responses toward the predicted values.

There is a large amount of variability noted in the response curves, especially in the groups that had series with long sequences. This is the case particularly in later stages of practice when it is not expected. One could speculate that many *Ss* "caught on" to the presence of long sequences early in learning and were becoming less motivated as the series progressed. Many *Ss* reported that they were "getting pretty bored" towards the end of the series and weren't "paying much attention to the task." They reported trying various repetitive responses (e.g., pushing the switch 10 times on one side, then 10 on the other) to break up the monotony. The *E*, also, noted many *Ss* dropping characteristic ways of responding and adopting different patterns of responding in the latter part of the series.

The analyses of the marginal data revealed that *Ss* overshot the more frequent stimuli or undershot the less frequent stimuli in the later stages of practice. Comparison of the present data with Engler's (1958) data showed that as the number of trials increased, *Ss* tended to go somewhat beyond the predicted number of marginal responses. It would seem that with sufficient trials and sequentially dependent reinforcements *Ss* tend not to match the marginal stimulus probabilities. It was also found that there were no differences in marginal response behavior due to the conditional stimulus probability specification.

Summary

This experiment on probability learning was designed to test several predictions from the Burke and Estes (1957) model and from an extension of this model by Witte (1959). A total of 120 *Ss* were tested on a Humphrey's-type light guessing apparatus. The series of reinforcing events were sequentially dependent.

The results showed that neither model adequately predicted performance on massed trials when the conditional stimulus probability was high, although both models did somewhat better at predicting performance when the conditional probability was low.

A prediction from the Witte model on performance on a series of distributed trials after a long series of massed trials was not upheld. There is some evidence that the model predicts qualitative but not quantitative changes from distributed trials.

The marginal response behavior revealed that response probability was consistently below the stimulus marginal probability.

References

Anderson, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

Burke, C. J., & Estes, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–146.

Engler, J. Marginal and conditional stimulus and response probabilities in verbal conditioning. *J. exp. Psychol.*, 1958, **55**, 303–317.

Estes, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154.

Hake, H. W., & Hyman, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

Rand Corporation. *A million random digits*. Glencoe, Ill.: Free Press, 1955.

Snedecor, G. W. *Statistical methods*. (4th ed.) Ames: Iowa State Coll. Press, 1946.

Walker, H. M., & Lev, J. *Statistical inference*. New York: Holt, 1953.

Wilcoxon, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid, 1949.

Witte, R. S. A stimulus-trace hypothesis for statistical learning theory. *J. exp. Psychol.*, 1959, **57**, 273–283.

(Received August 29, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 552–559

YOKED COMPARISONS OF INSTRUMENTAL AND CLASSICAL EYELID CONDITIONING[1]

JOHN W. MOORE AND I. GORMEZANO

Indiana University

The purpose of the present investigation was to compare the acquisition and extinction of CRs in classical and instrumental avoidance eyelid conditioning when the comparison was independent of the number and pattern of UCS occurrences. In addition, the study was concerned with assessing the effects of omitting versus delaying onset of the UCS on a fraction of trials, under the two conditioning procedures.

Eyelid conditioning studies by Logan (1951), Kimble, Mann, and Dufort (1955), and Hansche (1959), comparing classical and instrumental avoidance procedures have indicated a higher level of conditioning with the classical procedure. Logan (1959), and Kimble et al. (1955), have attributed the inferior level of conditioning of the instrumental procedure to its partial reinforcement schedule. Kimble et al., further suggested that the observed differences may reflect some nonassociative effects of the UCS (i.e., each UCS occurrence develops an increment of acquired drive) and not an associative or learning factor. Moreover, Spence (1958) assumes that drive level (D) is in part a function of the general level of emotionality resulting from previous administrations of a noxious UCS. Since in instrumental avoidance conditioning the CR prevents the occurrence of the UCS, reinforcement theory would seem to imply, therefore, that classical conditioning is superior because of the greater number of reinforcements and higher motivational level of S. Stimulus-response contiguity theory (Guthrie, 1952) would also predict superior conditioning under the classical procedure presumably, however, because it assures the occurrence of a vigorous response on every trial. On the other hand, statistical learning theory (Estes, 1959)[2] would attribute inferior conditioning of the instrumental procedure to the greater variability of stimulus conditions. These interpretations, based on previously reported studies, have attributed differences in conditioning level to differences in the number and pattern of UCS occurrences. Presumably then, if an experimental comparison were made in which the pattern and number of UCS occurrences (and their possible nonassociative effects) were the same, these interpretations would lead one to expect no differences in performance under the two conditioning procedures. The present investigation attempted to make this comparison employing two reinforcement techniques: Depending on response contingencies, the UCS was omitted on a fraction of the acquisition trials or merely delayed by an interval longer than that shown to produce little or no conditioning (McAllister, 1953).

METHOD

Apparatus.—The S's room was sound treated and contained two identical well

[1] This research was supported in part by a grant to the second author from the National Institute of Mental Health, Grant M-4034(A).

[2] W. K. Estes, personal communication, 1959.

illuminated enclosures permitting 2 *Ss* to be run simultaneously. The stimulating and recording components for the enclosures were identical. The enclosures were 3 ft. wide, 4 ft. deep, and 5 ft. high and their inside surfaces were painted a flat white. The *S* sat in the booth in a straight-backed chair and about 3 ft. in front of his eyes and slightly above his line of vision was a white rectangular box with a 10-cm. circular milk glass window. The CS consisted of an abrupt change in the illumination and hue (to reddish-orange) of this window upon discharge of a 2-w. General Electric NE 40 neon lamp. Also present in the enclosure was an intercommunication set and two 15-w. incandescent lamps which were kept lighted at all times to minimize beta responses (Grant, 1943; Grant & Norris, 1946). On top of the enclosures, and equidistant from their adjoining walls, rested an electric fan which provided a masking noise within the enclosures.

The *S* wore an elastic headband supporting a padded aluminum plate on which was mounted a rigid plastic stimulus air jet with a 2-mm. diameter orifice, and a microtorque potentiometer used to pick up *S*'s eyelid movement. Flexible plastic tubing from each of the air jets led to the two arms of a Y joint. The third arm of the Y joint was coupled to a rubber hose leading to *E*'s room and an ac-operated solenoid valve and tank of compressed nitrogen. The microtorque potentiometer was coupled to the right eyelid in a manner similar to that described by Grant, Schipper, and Ross (1952). A hypodermic tube was mechanically coupled to the shaft of the potentiometer and a length of piano wire was inserted in the tubing. To the other end of the piano wire was attached a small circular piece of thin aluminum. The aluminum tab was attached to *S*'s eyelid by adhesive tape, thus permitting movement of *S*'s eyelid to be transmitted to the shaft of the potentiometer. The frictional torque of the potentiometer was rated at about .004 in-oz. The signal from the potentiometer, generated by movement of the eyelid, was amplified by a Brush dc amplifier and recorded by means of an inkwriting Brush penmotor. Brush penmotors also recorded onset and termination of the CS and UCS.

Hunter-Brown and ATC electronic timers controlled durations of the CS, UCS, and CS-UCS interval. Activation of the timers and control of intertrial intervals was accomplished by a Western Union tape transmitter. The transmitter was pulsed by a cam-microswitch arrangement driven by a sychronous motor. The CS consisted of an abrupt change in illumination and hue of the milk glass window from .19 to .26 mL. which lasted for 600 or 3100 msec.[3] The UCS was a puff of compressed nitrogen of 100-msec. duration delivered to the right cornea and terminating with offset of the CS. The CS-UCS interval was either 500 or 3000 msec. The intensity of the air puff, measured at its point of delivery to the eye, was of sufficient intensity to support a 200-mm. column of mercury. The air puff was applied at a distance of ½ in. from the right eye of *S*, and eyelid movements were recorded from this eye.

A thyratron circuit, when switched on, operated a relay whenever an eyeblink of a preset sufficient magnitude occurred. The system was designed so that eyeblinks occurring 200 to 500 msec. after CS onset, could prevent occurrence of the UCS or delay its onset for 3000 msec. A switch permitted *E* to determine whether CRs from one or the other of the two recording systems would activate the circuit.

Procedure.—All *Ss* received 70 acquisition and 20 extinction trials. The intertrial intervals were 15, 25, and 35 sec., in a random order with an average of 25 sec. Eyelid closures of at least 1-mm. deflection from the baseline in the latency range of 200 to 525 msec. during acquisition and 200 to 600 msec. during extinction were considered to be CRs.

Experimental design.—The *Ss* were randomly assigned to five groups of 24 *Ss* with sex differences and any variability associated with the two recording systems balanced among groups. Four of these groups constituted the cells of a 2 × 2 factorial design in which instrumental avoidance and classical reinforcement conditions were made orthogonal to two partial reinforcement procedures. For an avoidance conditioning group (A) and avoid-yoke group (AY) 2 *Ss*, 1 from each group, were run simultaneously in a yoked-chair procedure in which a sufficient eyelid response in the 200 to 500-msec. interval after CS onset, by the *S* in Group A, precluded the occurrence of the UCS for both *Ss* on that trial. If the *S* in Group A failed to make a CR, the UCS was presented to both *Ss* 500 msec. after CS onset. An avoid-delay group (AD) and avoid-delay-yoke group (ADY) received an analogous

[3] Calibrations were made with a Spectra Brightness Spot meter manufactured by the Photo Research Corporation, Burbank, California. Readings were taken with a filter corresponding to the "red eye" of the CIE luminosity curve.

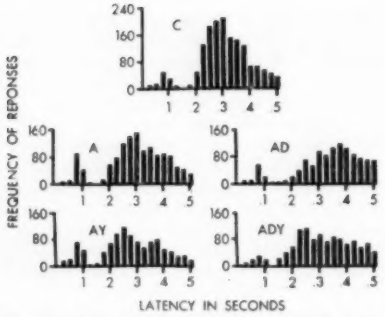


FIG. 1. Latency distributions of all eyelid responses in acquisition.

procedure in which a CR by the *S* in Group AD delayed the occurrence of the UCS for both *Ss* for 3000 msec. The fifth group, a conventional classical conditioning group (C), received the UCS on every trial at a CS-UCS interval of 500 msec. In extinction trials all groups received the CS alone for 600 msec.

Subjects.—The *Ss* were 120 men and women from introductory psychology courses at Indiana University. They were instructed to maintain a passive attitude toward the experiment and not to influence their natural reactions to the stimuli. Twenty-four *Ss* for each of the five conditions satisfactorily completed the experiment. In addition, 12 *Ss* were discarded from the experiment because of apparatus failure or experimental error.

RESULTS

The distribution of response latencies in acquisition for all five groups is shown in Fig. 1. The distributions are clearly bimodal with the first modes from the left showing the distribution of responses commonly taken as reflex responses to light (Grant, 1943). These responses are distributed about the 75-msec. mode with a latency range of approximately 40 to 120 msec. and occur with the highest frequency in Group A (avoid). The relatively low frequency of responses from approximately 100 to 200 msec., the beta range (Grant & Norris, 1946) suggests that precautions to maintain light adaptation were successful. Examination of the distribution of responses in the 200–525 msec. interval indicates some differences in the modal CRs among the five groups. The modal CRs for the two instrumental conditioning groups, A (avoid) and AD (avoid-delay), show slightly longer latencies than their yoked partial reinforcement groups, AY (avoid-yoke) and ADY (avoid-delay-yoke). Furthermore, Groups AD and ADY show a relatively greater frequency of CRs at the longer latency intervals. Inspection of records on delayed-UCS trials, however, indicated a relatively small number of responses in the 525–3000 msec. interval (not graphed). The mean number of eyeblinks in the 525–1025 msec. interval for Groups AD and ADY was 4.29 and 3.92, respectively, and in the 2500–3000 msec. interval, 8.04 and 5.25, respectively.

In Fig. 2 is shown the latency distribution of responses in extinction. The latencies of the modal CRs do not appear to have changed to any considerable extent from

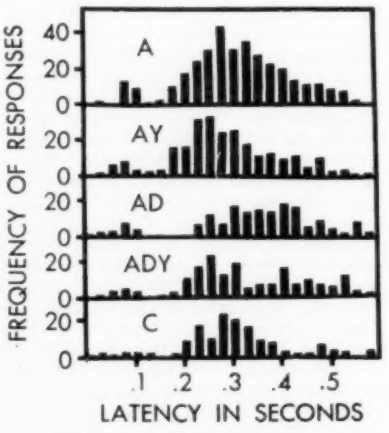


FIG. 2. Latency distributions of all eyelid responses in extinction.

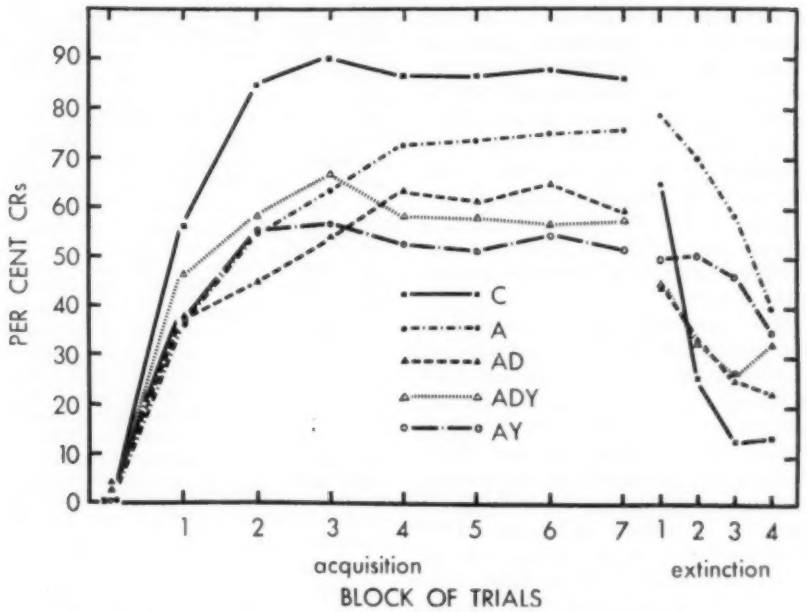


FIG. 3. The percentage CRs plotted in 10-trial blocks during acquisition and 5-trial blocks during extinction.

acquisition to extinction. There are, however, relative to Groups C, AD, and ADY, a greater number of responses elicited in Groups A and AY.

Figure 3 presents the results of plotting percentage CRs in acquisition and extinction. The initial points for all acquisition curves are the percentage responses on the first trial. The remaining acquisition points are the percentage CRs for each 10-trial block and points on the extinction curves are plottings of 5-trial blocks. Examination of the figure indicates that all groups perform essentially alike on the first trial. The rate of acquisition of CRs by Group C then proceeds to surpass that of the other groups, reaching a maximum of 90% on Trial Block 3 and stabilizing at around 87% for the remaining blocks of trials. Groups A and AY pursue essentially identical courses of acquisition up to Trial Block 2, where performance is at about 55%. Group AY remains approximately at this level while Group A continues its rapid upward course to 72% CRs on Trial Block 3 and from there proceeds to show a very gradual rise to a level of about 76% on the last 10-trial block. Since the UCS fails to occur in Groups A and AY when CRs are made by *Ss* in Group A, it is of interest to note that the stable performance of Group AY is being maintained under decreasing frequency of UCS occurrences. Also of interest, is the failure of Group A to reach asymptote. If further acquisition trials had been given it is conceivable that performance of this group might have reached the asymptotic level

of performance of Group C. Examination of the acquisition curves for Groups AD and ADY indicates that, for the first 30 trials, Group ADY is performing at a higher level. However, after Trial Block 3 there is a crossover in the two functions, with Group ADY dropping below Group AD to 58% CRs on Trial Block 4 while Group AD continues to rise until reaching an essentially asymptotic level of 63%.

The extinction curves presented on the right-hand side of Fig. 3 indicate a sharp drop in the percentage CRs in Group C, a lesser drop in Group A, and a brief rise followed by a drop in Group AY. Groups AD and ADY demonstrate essentially identical drops except that on the last five-trial block there is an upturn in Group ADY. The small drop in percentage CRs of Group AY relative to the large drop in Group A suggests that, under identical pattern and number of UCS occurrences, classically conditioned CRs (AY) may be more resistant to extinction than instrumentally conditioned CRs (A). A comparison of the extinction trend of Group C with the four other groups, looked at in terms of frequency of reinforcements (i.e., presentation of the UCS 500 msec. after CS onset), reveals what would be expected from the previous literature on partial reinforcement.

The percentage CRs on Trials 41–70 of acquisition and 1–20 of extinction of the five groups presented in Fig. 3 were treated by the arc-sine transformation to bring about homogeneity of variance (Bartlett, 1947). The two instrumental conditioning groups (A and AD) and their yoked classical conditioning groups (AY and ADY) were then subjected to an analysis of variance in which response contingency (instrumental and classical) was orthogonal to partial reinforcement technique (omission and delay of UCS). Table 1 presents the analysis of variance for the acquisition and extinction data. Except for the error mean squares which appear in parentheses at the bottom of each column, only the F ratios are presented. In Row 1, the source of variation tested is instrumental (A and AD) vs. classical (AY and ADY) conditioning. The F values indicate that the instrumental procedure was superior to the classical procedure at the .005 level for acquisition, but was not significant for extinction. Individual t test comparisons for acquisition yielded a P value of .01 for A vs. AY but failed to reach the .05 level for the AD vs. ADY comparisons. This indicates that the significant F value in favor of the instrumental procedure is primarily determined by the superior performance of Group A as compared with Group AY. Row 2 of this table shows that the partial reinforcement techniques of omitting vs. delaying the UCS (i.e., A, AY vs. AD, ADY) had no significant effect in acquisition, but produced an F value significant at the .01 level in extinction. Careful examination of the extinction curves in Fig. 3 indicates that this significant F was obtained because the UCS-omission groups (A and AY) elicited a greater number of responses in extinction than the UCS-delay groups (AD and ADY). Row 3 presents the source of variation attributable to sex and Row 4 to differences in the two recording systems. The F values for these sources of variation failed to reach significance in acquisition or extinction. Row 5 presents the interaction effects of response contingency (instrumental and classical) and partial reinforcement technique (UCS-omission and UCS-delay). The F for acquisition was significant at the .05 level, but failed to reach significance in extinction. Examination of the acquisition curves in Fig. 3 shows that this significant interaction effect is attributable to the fact that under the instrumental procedure the UCS-omission group (A) was superior to the UCS-delay group (AD), but under the classical procedure the UCS-delay group (ADY) was superior to the UCS-omission group (AY). The remaining sources of variation failed to reach statistical significance in acquisition, but in extinction Rows 6 and 7 resulted in F values significant at the .05 level. The source of variation tested in Row 6 is the interaction effects of sex and response contingency. Examination of percentage CRs in extinction indicates that under the instrumental procedures men gave an average of 32.62% and women 53.09% CRs, whereas, under the classical procedures men gave 38.31% and women 34.84% CRs.

Row 7 presents the interaction effects of recording system and sex. The significant F is due to the fact that under one recording system men gave an average of 31.98% and women 51.01% CRs while under the other recording system men gave 38.95% and women 36.92% CRs. Interpretation of these significant F values (extinction for Rows 6 and 7) would appear to be trivial and probably represent Type I errors.

Statistical comparisons in acquisition were made of the conventional classical conditioning group (C) with Groups A, AY, AD, and ADY and indicated that Group C was significantly superior to all other groups (t tests, 46 df, $P < .01$ in all cases except Group C vs. Group A, where $P = .05$). Similar comparisons in extinction revealed that Group A made significantly more CRs during extinction than the four other groups ($P < .01$, except A versus AY where $P = .05$); all other possible comparisons failed to show significant differences.

TABLE 1

ANALYSES OF VARIANCE OF PERCENTAGE CRs DURING ACQUISITION AND EXTINCTION

Source of Variation	F ratios		
	df	Acquisition	Extinction
Response contingency (RC): Instrumental (A and AD) vs. classical conditioning (AY and ADY)	1	9.05***	1.59
Partial reinforcement technique (PR): UCS-omission (A and AY) vs. UCS-delay (AD and ADY)	1	0.14	7.83**
Sex (S)	1	0.92	2.93
Recording system (RS)	1	3.70	0.51
RC × PR	1	5.09*	2.93
RC × S	1	1.15	5.81*
RS × S	1	3.77	4.49*
PR × S	1	1.06	1.87
PR × RS	1	0.46	0.06
RS × RC	1	0.00	0.01
RC × PR × S	1	1.83	1.71
RC × PR × RS	1	0.24	2.62
RC × S × RS	1	2.03	0.22
PR × S × RS	1	0.13	0.08
RC × PR × S × RS	1	0.45	1.61
Error	80	(383.61)	(592.18)

Note.—Number in parentheses are error mean squares. The analyses were performed on the arc-sine transforms of individual total percentage CRs for the last 30 acquisition trials and 20 extinction trials.

* $P = .05$.
** $P = .01$.
*** $P = .005$.

DISCUSSION

The principal findings of this investigation were: (*a*) the instrumental avoidance conditioning group (A), when compared with a classical conditioning group equated for number and pattern of UCS occurrences, demonstrated a superior level of conditioning; (*b*) the partial reinforcement procedure of UCS-omission resulted in greater resistance to extinction than UCS-delay; (*c*) a significant interaction was obtained in acquisition between response contingency and partial reinforcement technique i.e., under the instrumental procedure the UCS-omission group (A) was superior to the UCS-delay group (AD), but under the classical procedure the UCS-delay group (ADY) was superior to the UCS-omission group (AY); (*d*) the conventional classical conditioning group (C) was superior to all groups in acquisition but demonstrated the least resistance to extinction, whereas, the avoidance group (A) showed the greatest resistance to extinction.

The first finding is contrary to the speculations of Logan (1951) and Kimble, Mann, and Dufort (1955), suggesting that the inferiority of instrumental to classical eyelid conditioning could be accounted for in terms of partial reinforcement. The acquisition data of the present investigation indicate, in fact, that instrumental eyelid conditioning when equated in terms of partial reinforcement pattern and number of UCS occurrences, is superior to classical conditioning. This difference can not be adequately interpreted by present versions of reinforcement theory, contiguity theory, or statistical learning theory which have limited their formulations to the role of the UCS as a determiner of performance, without taking into account the relationship of the CR and occurrence of the UCS under the two conditioning procedures. In the instrumental procedure the UCS occurs only in the absence of the CR and in the classical procedure it occurs

independent of the CR. The superior performance of *Ss* in the instrumental procedure appears to be due then, to their receiving the UCS when *needed*, i.e., when the CR fails to occur, whereas, no such consistent relationship exists for the classical partial reinforcement procedure. If one were to assume that the association between the CS and CR is formed in an all-or-none fashion, as has been suggested by Estes (1960), then presentation of the UCS would have a maximum effect on response probability when the CR fails to occur and no effect when the CR does occur. Since in the yoked classical procedure any number of trials may elapse in which neither the CR or UCS occur it would be expected that this relationship would lead to performance that is inferior to the instrumental procedure.

The second and third finding bears on the function of the delayed-UCS procedure in classical and instrumental eyelid conditioning. Reynolds (1958), Ross (1959), and Ross and Spence (1960), in investigations of classical partial reinforcement, have employed the delayed-UCS procedure to control presumed motivational factors on nonreinforced trials. These experiments did not compare delayed-UCS partial reinforcement with the usual partial reinforcement technique of omitting the UCS completely on a certain percentage of trials. In the present investigation there were no significant differences in acquisition between the UCS-delay and UCS-omission procedures but UCS-omission did lead to greater resistance to extinction. The third finding suggests, however, that the two reinforcement techniques might operate differently under classical and instrumental conditioning. Under the instrumental procedure the UCS-delay technique (relative to UCS-omission) appears to attenuate performance, whereas, under the classical procedure it enhances performance. The exact function of UCS-delay remains equivocal however, since the present comparisons of UCS-delay groups with UCS-omission groups are necessarily limited by the fact that their respective schedules of reinforcement were generated by the performance of the two instrumental groups (A and AD), and these schedules were not the same. (The percentage reinforcement of Groups A-AY and AD-ADY were 39.58 and 52.13, respectively.)

The fourth finding indicates that acquisition of CRs in classical eyelid conditioning on a 100% reinforcement schedule is superior to instrumental and classical partial reinforcement conditioning even when care is taken to control possible nonassociative effects of the UCS (i.e. presentation of the UCS at the delay interval of 3000 msec.). The greater resistance to extinction of the avoidance group (A), in which the UCS did not occur on every trial, can perhaps be best accounted for in terms of Sheffield's (1949) stimulus trace hypothesis of partial reinforcement. According to this formulation, the perseverative stimulus traces of nonreinforcement become part of the CS complex operating during the following trial. These perseverating cues are precisely those which occur in extinction. More indirect support for Sheffield's hypothesis is provided by the similarity in extinction of the 100% reinforcement group with the UCS-delay groups. Since the occurrence of the UCS on all conditioning trials for the UCS-delay groups made the stimulus after effects following reinforced and nonreinforced trials highly similar, the stimulus trace hypothesis would predict little or no differences in extinction. This finding in support of Sheffield's hypothesis has the logical restriction however, of becoming definitive only if the null hypothesis can be proved.

Summary

The acquisition and extinction of CRs in classical and instrumental avoidance eyelid conditioning were compared when the comparison was independent of the number and pattern of UCS occurrences. In addition, the effects of omitting vs. delaying onset of the UCS on a fraction of trials were studied under the two conditioning procedures.

The results were as follows: (*a*) The instru-

mental avoidance conditioning group, when compared with a classical conditioning group equated for number and pattern of UCS occurrences, demonstrated a superior level of conditioning. (*b*) The partial reinforcement procedure of UCS-omission resulted in greater resistance to extinction than UCS-delay. (*c*) In acquisition the instrumental procedure of UCS-omission was superior to UCS-delay, but under the classical procedure UCS-delay was superior to UCS-omission. (*d*) A classical conditioning group on a 100% reinforcement schedule was superior to all groups in acquisition but demonstrated the least resistance to extinction, whereas, the instrumental avoidance group showed the greatest resistance to extinction.

These findings were discussed in terms of several contemporary hypotheses.

REFERENCES

BARTLETT, M. S. The use of transformations. *Biometrics*, 1947, **3**, 39–52.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

GRANT, D. A. Sensitization and association in eyelid conditioning. *J. exp. Psychol.*, 1943, **32**, 201–212.

GRANT, D. A., & NORRIS, E. B. Dark adaptation as a factor in the sensitization of the beta response of the eyelid to light. *J. exp. Psychol.*, 1946, **36**, 390–397.

GRANT, D. A., SCHIPPER, L. M., & ROSS, B. M. The effect of intertrial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44**, 203–210.

GUTHRIE, E. R. *The psychology of learning.* (Rev. ed.) New York: Harper, 1952.

HANSCHE, W. J. A comparison of operant, avoidance, and classical techniques in the conditioning of the eyelid response. Unpublished doctoral dissertation, University of Wisconsin, 1959.

KIMBLE, G. A., MANN, L. I., & DUFORT, R. H. Classical and instrumental eyelid conditioning. *J. exp. Psychol.*, 1955, **49**, 407–417.

LOGAN, F. A. A comparison of avoidance and nonavoidance eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 391–393.

MCALLISTER, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 417–422.

REYNOLDS, W. F. Acquisition and extinction of the conditioned eyelid response following partial and continuous reinforcement. *J. exp. Psychol.*, 1958, **55**, 335–341.

ROSS, L. E. The decremental effects of partial reinforcement during acquisition of the conditioned eyelid response. *J. exp. Psychol.*, 1959, **57**, 74–82.

ROSS, L. E., & SPENCE, K. W. Eyelid conditioning performance under partial reinforcement as a function of UCS intensity. *J. exp. Psychol.*, 1960, **59**, 379–382.

SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

(Received August 29, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 560-564

THE INTERTRIAL INTERVAL IN CLASSICAL CONDITIONING[1]

WILLIAM F. PROKASY AND FRANCIS L. WHALEY

Pennsylvania State University

The purpose of the present paper is to examine more closely the adequacy of the construct of response-produced inhibition as an explanatory principle for the common observation that performance under distributed trial conditions is superior to that under massed trial conditions. The principle is most concisely stated as Hull's (1952) Postulate IX:

> Whenever a reaction (R) is evoked from an organism there is left an increment of primary negative drive (I_R) which inhibits to a degree according to its magnitude the reaction potential ($_sE_R$) to that response. With the passage of time since its formation, I_R spontaneously dissipates approximately as a simple decay function of the time (t) elapsed . . . (p. 9).

One way in which this postulate can be tested is to compare the performance of independent groups of *S*s, each of which is conditioned with a different average intertrial interval. Three such studies (Baron, 1952; Prokasy, Grant, & Myers, 1958; Spence & Norris, 1950) have uniformly reported that response frequency in the classical eyeblink conditioning situation is positively related to the duration of intertrial interval.

An alternative test, employed in this study, can be effected by assessing response probability following individual intertrial intervals, with controls provided to balance out cumulative effects of the hypothesized inhibition. By training *S*s in a classical conditioning situation to frequency asymptote and employing extensive, balanced (but unpredictable) orders of intertrial intervals, it is possible to regroup the data of each *S* by separately evaluating response probability following each of the intervals employed. If Postulate IX is correct, response probability should increase as a function of the length of the specific interval separating trials.

[1] This research was supported by a National Science Foundation grant (G-7463) to the senior author.

METHOD

Apparatus.—The eyeblink conditioning apparatus employed was similar to that developed at the University of Iowa (Spence, 1953), as modified at the University of Wisconsin and described in Hansche and Grant (1960) and Prokasy, Grant, and Myers (1958). Eyelid movements were converted to changes in resistance by means of a piano wire attached to *S*'s right eyelid which in turn rotated the shaft of a microtorque potentiometer. Resistance changes from the potentiometer were electronically amplified and differentiated with both the response and its first time derivative recorded on a capillary pen ink-writing system with a paper drive operated at the rate of 150 mm/sec.

The *S*s were seated in a 3 × 5½ ft. room, facing a rectangular cubicle (34 × 34 × 28 in. deep) situated 31 in. above the floor. On the base of this cubicle facing *S* was a rectangular box, upon which were mounted two milk-white discs of 3 in. diameter. These were centered in the upper and lower halves of the CS box, with a signal light midway between them. Behind each milk-white disc was a 7-w. bulb, either of which could be employed as a CS. In this study only the upper light served in this capacity. The cubicle walls in *S*'s view were painted white, and were lighted by a 20-w. fluorescent lamp which was located out of *S*'s line of vision. This minimized the prospect of sensitized beta responses (Grant & Norris, 1947). The UCS, a puff of nitrogen, was delivered to the right cornea through a 5-mm. (inside diameter) glass jet.

The durations of CS and UCS signals were

controlled by Hunter-Brown interval timers. Trials were administered automatically by means of a taped program and a Western Union teletype transmitter. This permitted control of intertrial interval to within .25 sec.

Subjects.—The *S*s were 13 male students between the ages of 18 and 22 yr. They were volunteers from introductory psychology courses at Pennsylvania State University, and received course credit for their services. One of the 13 was terminated due to his failure to condition, apparently because of insensitivity to the puff stimulation.

Procedure.—Selected for investigation were the six intertrial intervals of 5, 10, 15, 20, 25, and 30 sec. These were chosen because prior data (Prokasy et al., 1958; Spence & Norris, 1950) indicated that performance differences were quite marked in the region of 15 to 30 sec.

Each *S* was given 97 reinforced trials in each of seven conditioning sessions. The first session was considered as a pretraining session, designed to bring *S* to asymptote. The remaining six sessions were test sessions, but data of only the final 78 trials in each of the test sessions were employed for analysis. By omitting the data of the first 19 test trials, the effects of any kind of recovery or regression were minimized. Thus, each *S* received a total of 679 trials, of which 468 were test trials employed for analysis.

In all sessions, including pretraining, all six intertrial intervals were administered in a predetermined, unpredictable order, with the restriction that there be a total of 16 trials following each interval length. (Note that this totals 96 intervals, which separate the 97 trials.) While the pretraining interval order was identical for all *S*s, a different order of intertrial intervals was administered for each of the six test sessions. Six different test-trial orders were employed, and were assigned to *S*s in a latin square design over the 6 test days. Thus, the overall design was a latin square with 2 *S*s in each of the six sequences of intertrial interval order. The only purpose in employing different sequences was the added assurance of being able to detect the existence, if any, of sequence-specific results.

The 78 test trials (i.e., those trials following the 78 test intervals) were ordered in such a way that each intertrial interval appeared 13 times, and that each of these intervals was preceded twice by all other intervals (including itself), during the session. This balancing was exercised in order to permit the assessment of possible residual effects from two trials preceding the test trial.[2] The net result was that, for each *S*, data pooled for the six test sessions provided a total of 78 trials immediately following each of the six intertrial intervals and 72 trials following each of the six intervals when they were two intervals prior to that trial.[3]

The CS was a brightness change of 3.8 to 190 mL. lasting 500 msec. The UCS, 50 msec. in duration, was directed at the right cornea from a distance of ½ to ¾ in., and flowed at the rate of 2 liters/min. (This rate was changed during the pretraining session for 1 *S* to a level of 2.75 liters/min. due to initial insensitivity.) The timed interstimulus interval was 470 msec., though the effective interstimulus interval was approximately 500 msec. This was determined by recording the transit time of the puff, approximately 30 msec. Because the six intertrial intervals were employed an equal number of times, the average intertrial interval was 17.5 sec.

Instructions were read to *S* at the beginning of the pretraining session. He was informed that he would experience two kinds of stimuli, a puff of air and a flash of light, and was instructed not to try to control his responses in any way. It was stressed that he should keep the CS panel in view at all times. Subsequent to the first day, instructions were not read, but *S*s were reminded, informally, of these instructions before each session. Due to difficulties in scheduling *S*s, it was not possible to run them on 7 consecutive days, though rarely was there a break between sessions of longer than 3 or 4 days.

Pen deflections of ½ mm. or greater were recorded as responses. While all responses

---

[2] The particular sequence is obtained easily by treating intertrial intervals as elements in a matrix. Form a symmetric 6 × 6 matrix with each row possessing all 6 elements. This will, when strung in a line, yield a system whereby each element is preceded by all other elements at least once. To insure that each element is preceded by itself requires the addition of, as a minimum, 3 elements prior to the appropriate rows, yielding a total of 39 elements. Perform the operation a second time and put the two strings together for a total of 78 elements.

[3] This differential comes about because certain intertrial intervals are employed in such a way that over the 6 test days there are six pairs of intertrial interval which constitute "extra" pairings. For purposes of analysis, the number of possible pairings had to be equal, hence, these "extras" were sampled out of the analysis.

TABLE 1

FREQUENCY OF RESPONSE FOR *S* DS AS A FUNCTION OF TEST DAY AND INTERTRIAL INTERVAL

Test Day	Intertrial Interval					
	5	10	15	20	25	30
1	10	11	11	9	11	11
2	10	9	11	10	13	8
3	11	11	11	9	11	10
4	12	10	12	10	10	9
5	13	9	10	11	12	12
6	12	11	10	12	8	11

were recorded, only the first response in the latency interval of 200 to 500 msec. was counted as a CR. The time derivative of the response was used as a more precise index of response latency. To insure that apparatus cues were not available to *S*, a masking noise of 80 db., presented through earphones, was on continuously during each session.

RESULTS

As a check for any unanticipated results inherent in test sequences over days, or for overall changes as a function of days, the total number of responses on each session was obtained for all *S*s and was employed as the dependent variable in an analysis of variance of the latin square design.

Based upon a residual variance of 484.11, the *F* ratios for day ($F = .1$; $df = 5/30$), sequence of intertrial interval orders ($F = 1.77$; $df = 5/30$), and intertrial interval order ($F = .07$; $df = 5/30$) were not significant. As anticipated, individual differences were significant ($F = 2.87$; $df = 6/30$). Because the overall 6×6 square was replicated, an assessment of square uniqueness was possible, and proved not to be significant ($F = 1.39$; $df = 20/30$). This result tends to confirm the assumption that those interactions, presumed to be negligible when latin square designs are employed, were negligible (Gourlay, 1955).

The basic analyses were performed on individual *S* data. In assessing the effects of intertrial interval on probability of response, a 6×6 matrix was formed. One dimension was sessions (confounded with specific intertrial interval order sequences) and the other dimension was intertrial interval. For purposes of illustration, a matrix of data for 1 *S* is provided in Table 1. An analysis of variance of response frequency as affected by sessions and intertrial interval was performed on such a matrix for each *S* separately. Because only a single value is present in each of the matrix cells, the error term employed was the interaction between session and intertrial interval. It was assumed, therefore, that this interaction effect was negligible, and to the extent that this assumption is not met the tests are conservative (Eisenhart, 1947).

The main effect of intertrial interval is significant for only 2 of the 12 *S*s. The data for all 12 are depicted in Fig. 1, those *S*s for whom the effect is significant being so indicated. Of immediate concern is the series of solid lines. Aside from two cases, the most striking feature is the zero slope of response frequency as a function of intertrial interval. This is a totally unanticipated result and it provides, insofar as negative results can be so interpreted, disconfirmation of the response produced inhibition hypothesis.

The results for the *S* (DW) whose data indicate a decreased response frequency as a function of intertrial interval are interpretable. This *S* had a tendency to doze momentarily between some trials during the last several minutes of the sessions. Subsequent examination of the records revealed that this was selective with the longer intervals. He was the only

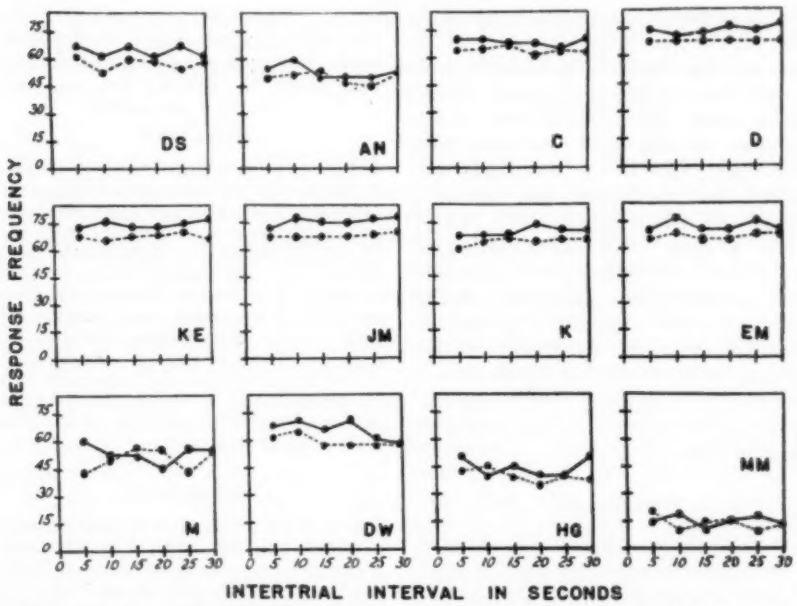


FIG. 1. Response frequency as a function of intertrial interval for each individual *S*. (Solid lines represent frequency following the immediate intertrial interval, dashed lines represent frequency on Trial *n* as a function of the interval preceding Trial *n*-1. Immediate intertrial interval effects are significant for *S*s M and DW.)

S in whom such behavior occurred. The second *S* (M) for whom the intertrial interval variable is significant was not consistent in his responding across this variable. We prefer to think that this is an instance of a Type I error.

The dashed line for each *S* in Fig. 1 depicts response frequency as a function of the intertrial interval *preceding* the interval immediately ahead of the test trial. Departure from slope zero would show residual aftereffects of the trial presented two trials before the test trial, and in no case was there significant departure.

An alternative index of cumulative effects was obtained by comparing response frequency following two successive 5-sec. intertrial intervals (5, 5) with response frequency following two successive 30-sec. intertrial intervals (30, 30). Of the 12 *S*s, 3 responded equally often following the 30, 30 and 5, 5 conditions, 1 responded more frequently following 30, 30, and 8 responded more frequently following 5, 5. Aside from ties, the likelihood by chance of 8 or more of 9 *S*s manifesting higher response frequencies to the same of two alternatives is .039 by a two tailed binomial test. This trend is in a direction *opposite* to that which would be predicted from the hypothesis of response-produced inhibition.

Analyses of response frequency as a function of days did not provide such unambiguous results. In 5 of the 12 *S*s, frequency varied significantly over days. In 2 cases, trends were manifest and explicable. In 1 of these, response frequency decreased systematically, accompanied by an

obvious decline in responsiveness to the UCS. In the other case, the *S* was one for whom UCS intensity was raised late in the pretraining series. As a result, the total number of responses on the first test day was depressed relative to those that followed, and this depression was sufficient to affect significantly the *F* ratio of days. For the remaining 3 *S*s whose performance over days was not constant, no systematic trends were observed. It is to be noted, however, that the likelihood of significant effects over days is increased because *E*, test day, test sequence, and time of day were confounded in this comparison.

Discussion

Because the present method permitted a test of Hull's Postulate IX with a range of intervals already established as extensively affecting performance, the negative findings are incompatible with theoretical expectation. Insofar as the classical conditioning situation is concerned, the hypothesis that the massed-distributed effects observed in other studies are the result of response-produced inhibition is inadequate to account for present data.

At the empirical level, the data suggest that a kind of "range effect" may be operating. Prokasy and Chambliss (1960) compared fixed intertrial intervals with varied intertrial intervals and found that group performance did not vary as a function of the amount of variation in intertrial interval. Together with the present study, this indicates that performance varies as a function of the *average* intertrial interval, but that there are no immediate differential effects for particular intervals within the range employed. That there is evidence of cumulative effects in a direction opposite to that expected on the basis of Postulate IX suggests that there may be certain relatively local variations in response frequency which are dependent upon the spread of intertrial intervals employed. Should further research verify this, the causative agent of the commonly observed massed-distributed effects would be anything but apparent.

Summary

Twelve *S*s were given extensive training in the classical eyeblink conditioning situation, where the main independent variable was the intertrial intervals of 5, 10, 15, 20, 25, and 30 sec. interspersed in an unpredictable order through the entire training series. Pooling the data of trials following each interval separately, it was found that response probability was not differentially affected by the immediate intertrial interval. Such results are incompatible with the hypothesis that response-produced inhibition accounts for the massed-spaced effects observed in earlier conditioning studies.

References

Baron, M. R. The effect of long intertrial intervals on the limit of eyelid conditioning. *J. exp. Psychol.*, 1952, **44**, 438–441.

Eisenhart, C. The assumptions underlying the analysis of variance. *Biometrics*, 1947, **3**, 1–21.

Gourlay, N. *F*-test bias for experimental designs of the latin square type. *Psychometrika*, 1955, **20**, 273–287.

Grant, D. A., & Norris, E. B. Eyelid conditioning as influenced by the presence of sensitized beta responses. *J. exp. Psychol.*, 1947, **37**, 423–433.

Hansche, W. J., & Grant, D. A. Onset versus termination of a stimulus as the CS in eyelid conditioning. *J. exp. Psychol.*, 1960, **59**, 19–26.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Prokasy, W. F., & Chambliss, D. J. Temporal conditioning: Negative results. *Psychol. Rep.*, 1960, **7**, 539–542.

Prokasy, W. F., Grant, D. A., & Myers, N. A. Eyelid conditioning as a function of unconditioned stimulus intensity and intertrial interval. *J. exp. Psychol.*, 1958, **55**, 242–246.

Spence, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, **45**, 1953, 57–63.

Spence, K. W., & Norris, E. B. Eyelid conditioning as a function of the intertrial interval. *J. exp. Psychol.*, 1950, **40**, 716–720.

(Received August 29, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 565–570

MEDIATION IN PAIRED-ASSOCIATE LEARNING

NAN E. McGEHEE AND RUDOLPH W. SCHULZ

Northwestern University *University of Iowa*

Russell and Storms (1955) have shown that language habits inferred from free-association norms are presumably capable of mediating the learning of verbal paired associates. Their *S*s learned to associate a word (B) with a nonsense syllable (A) in learning the first of two lists. The response unit B was the first word of an associative-chain (B-C-D) inferred from the norms. Following A-B learning, *S*s learned a test list consisting of A-D and A-X pairs. The A-D pairs were learned significantly faster than the A-X control pairs. However, as Russell and Storms (1955) hasten to point out, "The mere demonstration of mediational influences in learning . . . does not explain how the effect is achieved" (p. 292). One main purpose of the present experiments was to attempt to provide this explanation for the above situation.

As pointed out recently (Underwood & Schulz, 1960), it is analytically fruitful to conceive of verbal learning as a two-phase process. The first phase consists of a response-acquisition or response-recall phase where *S* is concerned with learning or recalling the response units per se. The second phase, the associative phase, involves learning to associate the response units with their appropriate stimulus units.

The usual interpretation of the Russell and Storms (1955) result has been the one that the associative phase of test-list acquisition was facilitated via the specific associative chains linking the A and D items of the respective pairs. An alternative interpretation, made apparent by the two-phase conception, is that facilitation occurred because the B items on the first list enhanced the availability of the D items during A-B learning by raising them in *S*'s response hierarchy. Moreover, this enhanced availability could facilitate test-list acquisition in the absence of any specific associative link between A and D. That is, the increased ease of response recall which would accompany heightened availability should facilitate test-list acquisition. The latter expectation is consistent with the results of a recent study in which it was found that the acquisition of paired-adjective lists was facilitated when the availability of the response units was enhanced by deliberate pretraining (Underwood, Runquist, & Schulz, 1959). The availability hypothesis was tested in Exp. I.

A second purpose of Exp. I was to compare performance under Russell and Storms' (1955) mediated and nonmediated conditions with performance under a condition in which first and test lists were unrelated (i.e., the practice control of transfer experiments). This was done to determine, in somewhat more absolute terms, the amount of facilitation which results from mediation in this situation.

The third purpose of Exp. I was to extend the generality of the Russell and Storms' (1955) results by replicating them with a design consisting of independent random groups and homogeneous lists as well as including *S*s of both sexes.

The results of Exp. I failed to

support the availability hypothesis. However, for reasons too lengthy to detail here, we remained skeptical as to whether the failure of the availability hypothesis necessarily implied that mediation had taken place via specific associative chains. Therefore, Exp. II was undertaken to investigate the matter further. Namely, if there is specific linkage, then a test list in which the A and D items are inappropriately paired (i.e., the analogue of the S_1-R_1, S_1-R_r paradigm of conventional transfer terminology) should be more difficult to learn than the nonmediated test list. Put another way, there should be mediated interference. Indeed, Norcross and Spiker (1958) have demonstrated just such mediated interference for associative links acquired entirely in the laboratory. In Exp. II an attempt was made to demonstrate mediated interference in the present situation.

A second purpose of Exp. II was that of determining the "criticalness" of the free-association norms as predictors of the mediation effects obtained in Exp. I. That is, while the B and D items were selected so as to minimize the possibility of a *direct* free-associative link between them, the method of selection *did not* preclude the possibility of a meaningful relationship between these items (e.g., Thief-Take, Wish-Need, etc.). Similarly, it did not prevent the occurrence of interlist relationships between response units in terms of formal similarity (e.g., Smell-Stem, Memory-Matter, etc.). Hence, if the interlist relationships between response units, along dimensions *not* predictable from free-association norms, were stronger in the mediated condition than in the nonmediated condition, then the interpretation of the superior performance in the mediated condition would require revision. Therefore, Exp. II included a condition in which the first and test lists were learned in reverse order (A-D, A-B). The logic of this arrangement is based on the assumption that the associative linkage defined by the free-association norms is unidirectional (i.e., B-C-D, but not D-C-B). Thus, making A-D the first list and A-B the test list, should reduce substantially, or even eliminate, the facilitation presumed to be mediated by habits inferred from free-association norms. Contrariwise, a relationship between B and D based on interlist response similarity should be unaffected by the reversal of the lists because it would be expected to be a bidirectional relationship. Storms (1958) has proposed and tested a similar hypothesis for a situation involving a single mediating term. He found clear evidence for bidirectionality with normatively unidirectional materials.

METHOD

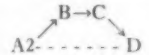
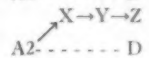
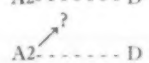
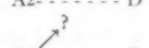
Lists: Exp. I.—The relationships between first and test lists define the conditions of Exp. I, and are summarized in Table 1. The first-list designations shown in Table 1 will be used as abbreviations in subsequent references to the various conditions (e.g., the mediated condition will be called Cond. A2-B, the nonmediated condition will be Cond. A2-X, etc.).

The stimulus designations A1 and A2 are used to distinguish between the two sets of 10 nonsense syllables required by the design of Exp. I. The syllables were selected so that inter- and intralist similarity would be at a minimum. Their association values ranged from 0% to 27% according to Glaze (Underwood & Schulz, 1960, Appendix A). In minimizing similarity it was impossible, with one exception, to retain the syllables used by Russell and Storms.

The B, D, and X response units were the same words as those used by Russell and Storms (1955) and are shown in Table 3 of their article (p. 290). The details regarding the selection of these words may also be found there. In brief, C was the most frequent free-association response to B, and D the most

TABLE 1

RELATIONSHIPS BETWEEN FIRST AND TEST LISTS DEFINING CONDITIONS OF EXP. I

Condition	First List	Mediation Chain	Test List	Inferred Action
Mediated	A2-B	B-C-D	A2-D	
Nonmediated	A2-X	X-Y-Z	A2-D	
Availability	A1-B	B-C-D	A2-D	
Practice control	A1-X	X-Y-Z	A2-D	

frequent free-associate of C, but D was not among the 10 most frequent responses to B, e.g., B (Trouble), C (Bad), D (Good). The members of the X-Y-Z chain are related to one another in the same manner as the members of the B-C-D chain; however, the members of the two respective chains were not related, and must not be related, to each other. Each list consisted of 10 paired associates.

Conditions A2-B and A2-X correspond to Russell and Storms' (1955) chained (A-D) and unchained (A-X) conditions, respectively. The availability hypothesis would be supported if test-list performance in Cond. A1-B is superior to that in Cond. A1-X. Similarly, comparison of A2-B with A1-X will determine if absolute positive transfer resulted from mediation.

Lists: Exp. II.—Conditions A2-B and A2-X of Exp. I were replicated. The lists of Cond. A2-D were the same as those of Cond. A2-B in Table 1 except that A2-D was the first list and A2-B the test list. Condition A2-Br is most easily described as follows. Let A_1-B_1, A_2-B_2, etc. represent the syllable-word pairings of the first list. Similarly, B_1 is the beginning of the free-association chain B_1-C_1-D_1. The test list represents a random re-pairing of the respective A and D items: A_1-D_4, A_2-D_{10}, A_3-D_1, etc. This test list was List A2-D—the same one used in Exp. I and in Cond. A2-B and A2-X of this study.

Mediated interference will be demonstrated if performance in Cond. A2-Br is inferior to that in Cond. A2-X. If mediation in these materials is unidirectional, performance in Cond. A2-D and A2-X should not differ.

Procedure: Exp. I and II.—Each *S* was read standard instructions for paired-associate learning prior to learning the first list. This list was then learned to a criterion of three consecutive errorless recitations. Following the completion of first-list learning all *S*s rested for 4 min. After brief instructions to proceed as before, *S*s were presented the test list for 10 anticipation trials. For those *S*s not reaching a criterion of one errorless recitation during the first 10 trials, test-list acquisition was continued until this criterion was reached. The lists were presented on a memory drum at a 2:2-sec. rate with a 4-sec. intertrial interval. The 10 pairs in each list were presented in five random orders to prevent serial learning of the response units. The experimental session was limited to 50 min. for all *S*s.

Subjects.—The *S*s, Northwestern University undergraduates, were randomly assigned to conditions, with 30 and 24 *S*s per condition in Exp. I and II, respectively. The *S*s were naive with respect to the materials used, although most of them had served in other verbal learning experiments prior to their present service. When an *S* did not complete the experiment he was replaced by the next *S* appearing at the laboratory. There was no relationship between failure to complete the experiment and conditions.

RESULTS AND DISCUSSION

First-list acquisition: Exp. I and II.—Performance, in terms of mean number of trials to reach the criterion of three consecutive perfect recitations, did not differ significantly under the four respective conditions

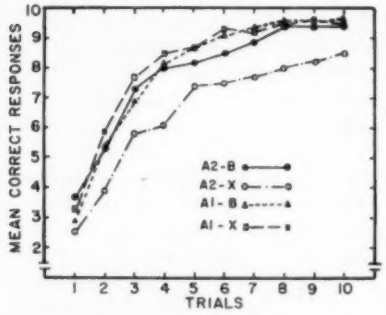


FIG. 1. Test-list performance in Exp. I as a function of various relationships between the stimulus and response units of the first and test lists. (See text for a complete description of these relationships.)

of either experiment (Exp. I: $F = 1.77$, $P > .10$; Exp. II: $F < 1$). In Exp. I the means were 18.33, 19.23, 18.90, and 15.43 for Cond. A2-B, A2-X, A1-B, and A1-X, respectively. The respective means for Cond. A2-B, A2-D, A2-X, and A2-Br of Exp. II were 13.62, 14.75, 14.71, and 15.87. The comparable difficulty of Lists A2-B and A2-D as first lists permits direct comparisons of test-list performance under Cond. A2-D with the other conditions of Exp. II, since the test list for this condition was List A2-B while List A2-D served as test list for the other conditions.

Test-list acquisition: *Exp. I.*—Test list performance is shown in Fig. 1. It is apparent from Fig. 1 that performance under Cond. A2-B, A1-B, and A1-X was essentially equivalent. Performance under Cond. A2-X was consistently inferior to performance under the other three conditions. The mean total number of correct responses during Trials 1–10 was 78.23, 65.43, 79.50, and 81.30 for Cond. A2-B, A2-X, A1-B, and A1-X, respectively. The standard error of these means ranged from 1.64 to 2.75. The only reliable ($P < .01$) differences among these means are those involving a comparison between Cond. A2-X and each of the other three conditions. The results for mean number of trials to reach the criterion of one perfect recitation were in complete agreement with those for total correct responses.

From these results it seems clear that Russell and Storms' (1955) findings were reliable and of some generality. The significantly ($t = 3.79$, $P < .01$) superior performance under mediated as contrasted with nonmediated conditions in the present study represents a reproduction of their results. The interlist relationship in these two conditions—S_1-R_1, S_1-R_2—is such that negative transfer would ordinarily be expected. Hence, the failure to find negative transfer under the mediated condition and the presence of a substantial amount of negative transfer in the nonmediated condition indicates that the inhibitory effects of the interlist relationship were somehow overcome in the mediated condition. However, comparison of performance under mediated and practice-control conditions makes it clear that the facilitation produced by mediation was not of sufficient magnitude to produce absolute positive transfer (see Fig. 1).

Finally, it is apparent that the avail-

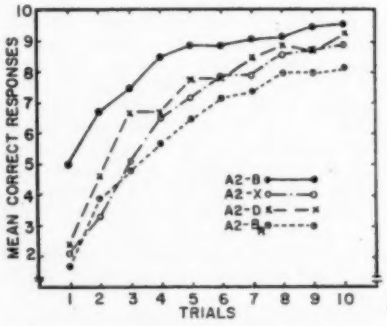


FIG. 2. Test-list performance in Exp. II as a function of various relationships between the stimulus and response units of the first and test lists. (See text for a complete description of these relationships.)

ability hypothesis was not supported by the results of the present study. That is, had the availability of the D units of the test list in Cond. A1-B been enhanced during first-list acquisition, then test-list performance under Cond. A1-B should have been superior to performance under Cond. A1-X. As can be seen from Fig. 1, it was not. Moreover, an analysis of overt errors failed to adduce any evidence indicating enhanced response availability in Cond. A1-B. Therefore, mediation effects, at least in the present situation, cannot be attributed to facilitation of the response-recall phase of test-list acquisition via enhanced response availability.

Test-list acquisition: Exp. II.—The performance on the test list for Trials 1–10 is shown in Fig. 2. The means for the total number of correct responses over the 10 trials were 83.04, 66.37, 71.62, and 61.54 for Cond. A2-B, A2-X, A2-D, and A2-Br, respectively. The standard error of these means ranged from 1.75 to 2.57. The overall differences among means are highly reliable ($F = 10.55$, $P < .01$).

Significant ($t = 5.12$, $P < .01$) superiority of performance in Cond. A2-B over Cond. A2-X again replicates Russell and Storms' (1955) results along with those of Exp. I.

As can be seen from Fig. 2, performance was initially comparable in Cond. A2-X and A2-Br, but after Trial 3 the curves diverge with performance under Cond. A2-X remaining consistently superior on Trials 4 through 10. The conditions did not differ significantly ($t = 1.45$, $P > .10$) in terms of total correct over Trials 1–10. However, the difference in mean performance is clearly in the expected direction. Furthermore, consistent with the divergence in the curves of Fig. 2 for these conditions, the means for the number of trials to reach the criterion of one perfect recitation were 8.92 for Cond. A2-X and 11.87 for Cond. A2-Br. This difference in trials to reach criterion is highly significant ($t = 2.65$, $P < .01$). It appears justifiable to conclude that mediated interference has been demonstrated in Cond. A2-Br. This result agrees with the one obtained by Norcross and Spiker (1958) with associative chains learned in the laboratory.

Inspection of Fig. 2 reveals a trend toward slightly better performance, at least initially, in Cond. A2-D than in Cond. A2-X. However, when the means for total correct on the 10 test-list trials for Cond. A2-D and A2-X were compared they failed to differ significantly ($t = 1.61$, $P > .10$). Indeed, in terms of mean trials to criterion Cond. A2-D was slightly inferior to Cond. A2-X. Furthermore, performance under Cond. A2-D was reliably poorer ($t = 3.51$, $P < .01$) than under Cond. A2-B.

Thus interlist response similarity along dimensions not predictable from free-association norms does not appear to be a major contributor to the facilitation of test-list performance in the present situation. This result appears to conflict with the one described earlier, which Storms (1958) obtained. However, it should be recalled that his situation differed from the present one in that it involved only a single mediating term in contrast to the present two term mediating link. Perhaps, the potency of the "recency" effect to which he attributes his results varies inversely with the length of mediating chain. If this is the case, the present contradiction would be resolved easily.

In short, the results of Exp. I and II seem to provide fairly conclusive evidence that mediation in the Russell and Storms (1955) situation can be explained in terms of the facilitation of the associative phase of test-list acquisition via specific unidirectional associative chains linking the respective A and D items of the test-list pairs. Moreover, this inter-

list response relationship is associated with previously acquired language habits reflected by free-association norms.

Summary

Two experiments were conducted to determine, among other things, how language habits inferred from free-association norms might mediate the learning of verbal paired associates in the Russell and Storms' (1955) situation. Their general procedure was replicated except that independent random groups and homogeneous lists as well as *Ss* of both sexes were used. The interlist relationship between the stimulus syllables and response words of the first and test lists were appropriately varied to define the various conditions under which the 216 *Ss* of the present experiments learned the 10-item lists on a memory drum at a 2:2-sec. rate.

From the results it was concluded that: (*a*) The response-recall phase of test-list acquisition is not facilitated in the mediated condition. (*b*) Mediated interference can be produced with the Russell and Storms' (1955) materials. (*c*) Under the mediated condition, the associative phase of test-list acquisition is facilitated via the specific associative chains linking the stimulus and response units of the respective pairs. (*d*) The free-association norms are "critical" in defining these associative chains. (*e*) Russell and Storms' (1955) results are reproducible and of considerable generality even though the facilitation produced by mediation does not result in absolute positive transfer.

References

Norcross, K. J., & Spiker, C. C. Effects of mediated associations on transfer in paired-associate learning. *J. exp. Psychol.*, 1958, **55**, 129–133.

Russell, W. A., & Storms, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

Storms, L. H. Apparent backward association: A situational effect. *J. exp. Psychol.*, 1958, **55**, 390–395.

Underwood, B. J., Runquist, W. N., & Schulz, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received September 2, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 571–575

STUDIES OF DISTRIBUTED PRACTICE: XXI. EFFECT OF INTERFERENCE FROM LANGUAGE HABITS[1]

BENTON J. UNDERWOOD AND RUDOLPH W. SCHULZ

Northwestern University *State University of Iowa*

The last report in this series (Underwood & Schulz, 1961) indicated that when distributed practice facilitates acquisition of paired-associate lists it does so as a consequence of interference occurring during the response-integration phase of learning. The interference was produced by having *Ss* learn lists containing elements which subsequently interfered with the response learning of the critical or test list. The present experiments continue the exploration of the interaction between intertrial interval and interference in verbal learning. However, the interference introduced into the lists results primarily from language habits brought to the laboratory by *Ss*. Two types of experiments will be reported, distinguished by the locus of the interference. In the first type, interference is localized in the response terms and is presumed to hinder the acquisition of responses as such. In the second type, familiar words are used so that there should be no problem of response integration. Rather, the interference comes about as *S* attempts to attach particular responses to particular stimuli. If the previous findings are generalizable, facilitation by distributed practice may be expected with some level of interference as produced by the first type of experiment, but no facilitation should be expected in the second type of experiment.

[1] This work was done under Contract Nonr-1228(15), Project 154-057, between Northwestern University and the Office of Naval Research. We are grateful to Geoffrey Keppel, who supervised the gathering of the data.

METHOD

Three experiments will be reported. The procedures common to all experiments are as follows: (*a*) For distributed practice (DP) the intertrial interval was introduced after each trial and was always 30 sec. in length. Symbol cancellation was used to fill this interval. For massed practice (MP) the intertrial interval was 4 sec. (*b*) All lists were paired-associate lists. The rate of presentation was 2 sec. for the stimulus alone (anticipation learning), and 2 sec. for the stimulus and response together. Four different orders of presenting the lists on the memory drum were employed. (*c*) In a given experiment, no *S* served in more than one condition. Most *Ss*, all undergraduate students at Northwestern University, were naive to formal verbal-learning experiments; if not, *S* was used only if the materials learned previously could not, in our judgment, interfere with the materials to be learned in the present experiments. (*d*) For each experiment a randomized order of conditions was made up and *Ss* were assigned to this order as they appeared at the laboratory. (*e*) All lists were presented for a constant number of trials with *Ss* instructed to give as many correct responses as possible on each trial.

Experiment I.—Three sets of eight trigrams were made up such that the average associative connection between letters may be spoken of as low, medium, and high. The associative connection was determined from the Underwood-Schulz (1960) tables. Within each set of eight trigrams formal intralist similarity was zero in that no letters were repeated. These trigrams were used as response terms with single-digit numbers as stimuli. Since each list was presented under MP and under DP, six groups of *Ss* were required. Each of these groups consisted of 28 *Ss*; all were given 20 anticipation trials.

Experiment II.—Two sets of eight bigrams each were constructed so that the average associative connection between the two letters for one list was low, for the other, high. The bigrams having low integration were: EY, CF, XV, DS, QW, JH, MK, and RZ. The bigrams having high integration were: MN, RA, CO, QU, XY, ET, JK, and

TABLE 1

PAIRED-ASSOCIATE LISTS USED IN EXP. III

List 1	List 2	List 3	List 4	List 5
CRUISER-HEAD	HEAD-MEASLES	BOB-HEAD	BOB-MEASLES	BOB-NITROGEN
EMERALD-WALL	BUS-MUMPS	BILL-BUS	BILL-MUMPS	BILL-SPARROW
BLUE-COTTON	KNIFE-POLIO	JOE-KNIFE	JOE-POLIO	JOE-GERMANY
THEFT-BEE	JOHN-CANCER	JOHN-CANCER	JOHN-CANCER	JOHN-CANCER
DOCTOR-BUS	WALL-RUSSIA	COW-WALL	COW-RUSSIA	COW-BLUEJAY
COPPER-TABLE	DOG-FRANCE	DOG-FRANCE	DOG-FRANCE	DOG-FRANCE
DOG-FRANCE	TROUT-ENGLAND	CAT-TROUT	CAT-ENGLAND	CAT-HYDROGEN
GASOLINE-MAPLE	DAISY-GERMANY	HORSE-DAISY	HORSE-GERMANY	HORSE-POLIO
WALTZ-OXYGEN	COTTON-BLUEJAY	RABBI-COTTON	RABBI-BLUEJAY	RABBI-MEASLES
WINE-GEOLOGY	MAPLE-CANARY	BISHOP-MAPLE	BISHOP-CANARY	BISHOP-RUSSIA
TRUMPET-KNIFE	GEOLOGY-SPARROW	MINISTER-GEOLOGY	MINISTER-SPARROW	MINISTER-SULPHUR
APPLE-TROUT	PRIEST-ROBIN	PRIEST-ROBIN	PRIEST-ROBIN	PRIEST-ROBIN
HAT-SECOND	BEE-NITROGEN	RHUMBA-BEE	RHUMBA-NITROGEN	RHUMBA-MUMPS
JOHN-CANCER	TABLE-HYDROGEN	FOXTROT-TABLE	FOXTROT-HYDROGEN	FOXTROT-ENGLAND
FOOTBALL-DAISY	WALTZ-OXYGEN	WALTZ-OXYGEN	WALTZ-OXYGEN	WALTZ-OXYGEN
PRIEST-ROBIN	SECOND-SULPHUR	TANGO-SECOND	TANGO-SULPHUR	TANGO-CANARY

DI. Single-digit numbers were again used as stimulus terms and acquisition was carried for 15 trials. Four groups of 15 *Ss* were employed (MP and DP for each list). Subsequently, a partial replication was carried out by adding 15 *Ss* to the two groups assigned the low-integrated list.

Experiment III.—The above two experiments were concerned with the effects of DP with varying degrees of response integration. In Exp. III the responses were common words. These words were taken from the norms given by Cohen, Bousfield, and Whitmarsh (1957). Their *Ss* (a total of 400) were given category names under instructions to write down the first four instances of the category that occurred to them. The tables presented by the investigators are in terms of the frequency with which each response was given. For example, to the category An Article of Furniture, 392 of the 400 *Ss* gave CHAIR, 346 TABLE, and so on. The nature of the lists constructed from such materials can best be understood by examining the lists. These are presented in Table 1.

First, attention is directed to List 2. In this list there are four sets of response words, of four each. The sets are distinguished in terms of category words. That is, MEASLES, MUMPS, POLIO, and CANCER represent responses given to the category stimulus "A Disease." The words RUSSIA, FRANCE, ENGLAND, and GERMANY are responses to the stimulus category A Country. The four words representing each category are words given with high frequency to the category stimulus. The assumption made was that because these items are all connected to a common category name (a common mediator), or because they have high associative connection among themselves as a consequence of cultural usage or as a consequence of being associated with a common mediator, differences in learning might be expected as a consequence of certain manipulations of these clusters. With these notions in mind, an examination of the structure of each list may be made.

List 1. There are no clusters of words from the same category.

List 2. Four clusters of four items each occur on the response side, with no clusters among the stimulus terms.

List 3. The reverse of List 2. That is, there are four clusters among the stimulus terms and none among the response terms.

List 4. Four clusters on both the stimulus and response side and the pairing of stimulus and response terms is such that all items of a given cluster on the stimulus side are paired with all items of a given cluster on the response side; e.g., animals are always paired with countries.

List 5. Four clusters on both the stimulus and response sides, but the pairing is such that the four members of a cluster on the stimulus side are paired with items from four different clusters on the response side.

In presenting the lists the order of the pairs was essentially random and four different orders were used. Each list was presented for 15 trials. A total of 150 *Ss* was employed, divided into 10 groups of 15 each. Five groups were given the lists by DP and five by MP.

RESULTS

Experiment I.—The response measure was the mean total number of

correct responses given over 20 trials. The results are shown in Fig. 1. The statistical analysis shows that only the F for response integration was significant; this effect is largely produced by differences between the low and medium list. The F for MP vs. DP was less than 1. Clearly, there is no evidence in these data that DP facilitates acquisition at any level of integration of these trigram responses.

Experiment II.—The response measure used was the mean total number of correct responses over 15 trials. For the bigram list with low integration the mean number of total correct responses given over 15 trials under MP was 47.4, and under DP, 62.5. For the list with high integration the means were 88.5 for MP, and 87.3 for DP. The F for interaction is 4.75, with 3.96 needed for the 5% significance level for 1 and 86 df. The t between MP and DP for the list with low integration is 2.55 when based on the variance of these two groups only, the σ_{diff} being 5.93. The acquisition curves over the 15 trials showed that DP was superior to MP throughout. Finally, it will be remembered that to the original 15 *S*s assigned each group, 15 were added to the two groups learning the list of low integration. For the original two groups assigned to this list the difference between MP and DP in total correct responses was 15.8; for the two added groups the difference was 14.4. We conclude that DP facilitates acquisition of bigrams of low integration and does not facilitate a list of well-integrated bigrams.

The learning of lists with response trigrams of low integration was not facilitated by a DP interval of 30 sec.; the learning of bigram lists of low integration was facilitated. As will be noted later, we are inclined to believe that there is a critical interaction between degree of response integration and length of intertrial interval. However, it is possible that in the case of bigrams the frequency of second letters is the critical factor. Both lists of bigrams had the same first letters, but to obtain low associative connection between the two letters the second letter must usually be of low frequency and to obtain high associative connection the second letter must usually be of high frequency. It is conceivable, therefore, that frequency of second letters is the critical difference between the two lists and not degree of integration. An indirect test was made of this notion by an experiment in which only single letters were used as response terms, one list having letters of low frequency (Q, Z, J, X, K, V, B, Y) and another having letters of high frequency (A, T, N, S, O, I, R, E). Again using single-digit numbers as stimuli, the lists were presented under MP and under DP. For both lists the means were almost identical for MP and DP. It may be concluded that it is not simply the frequency of second letters of the bigrams which are critical for the facilitation by DP found with the

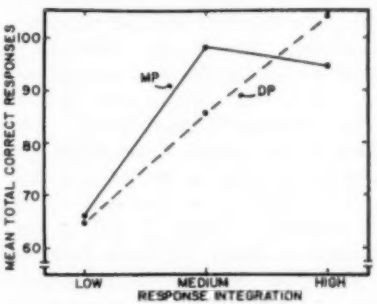


FIG. 1. Paired-associate learning as a function of degree of response integration of trigrams and massed vs. distributed practice.

TABLE 2

MEANS AND *SD*s OF TOTAL CORRECT RESPONSES IN LEARNING LISTS OF EXP. III AS A FUNCTION OF MASSED AND DISTRIBUTED PRACTICE

List	Massed Practice		Distributed Practice	
	Mean	*SD*	Mean	*SD*
1	197.40	22.70	197.33	24.18
2	189.80	29.83	175.27	32.35
3	176.13	35.74	187.60	21.85
4	157.33	35.47	143.40	36.24
5	159.20	24.27	147.73	28.96

list of low integration in Exp. II.

Experiment III.—The mean total numbers of correct responses given in 15 trials are shown in Table 2. An analysis of variance shows that variance attributable to list differences is highly significant statistically ($F = 13.20$, $df = 4/140$, $P < .01$). The F for MP-DP is 1.30. Not only does this fall far short of significance, but it may be noted that the direction of the difference in learning tends to to favor MP.

That differential interference is responsible for the differences in learning is suggested by the frequency of overt errors. These frequencies are, in general, inversely related to amount learned over the 15 trials and are not related to intertrial interval. For the MP and DP conditions combined, the mean numbers of total errors made in learning are: 8.8, 20.5, 14.9, 47.8, and 28.0 for Lists 1 through 5, respectively. That there is not a perfect correspondence between number of errors and slowness in learning is apparent in two comparisons. Lists 2 and 3 were learned at about the same rate, but more errors were made in learning List 2. The fact that more errors were made when the concepts were on the response side than when on the stimulus side corresponds to the results of earlier studies (Underwood, 1953) where clusters of highly similar adjectives were used. The second comparison is that between Lists 4 and 5. Again, the rate of learning was approximately the same for these two lists but the number of errors is much greater for List 4 than for List 5. All of these differences in error frequency maintain themselves even if adjustments are made for differences in opportunity to make errors because of differences in learning rate.

DISCUSSION

Considered in conjunction with previous experiments, the present three studies support two generalizations: (*a*) DP will facilitate acquisition only when interference occurs in integrating response terms, and (*b*) DP will not facilitate acquisition when interference, no matter how heavy, occurs in associating the stimulus and response terms of a paired-associate list. Obviously the first generalization must be limited in the sense that DP will not inevitably facilitate acquisition when interference occurs in acquiring response terms. In the present studies DP did not facilitate acquisition of trigrams at any level of integration; DP facilitated only the list of low-integrated bigrams. Clearly, a further narrowing of conditions under which DP will facilitate is required. Two statements will summarize our present position on this matter. First, some minimal amount of interference in response acquisition is necessary before DP will facilitate learning. Second, as amount of interference increases beyond this minimal amount, there will be a sharp interaction between length of the DP interval and the amount of interference. More specifically, it is believed that the greater the amount of interference the shorter must be the DP interval for facilitation to occur. The full elaboration of this position, along with some resolution of apparent contradictions to it now existing in the

literature, has been made elsewhere (Underwood, 1961).

Summary

Three studies evaluated the influence on learning of the interaction between intertrial interval and interference. The interference in Exp. I and II was centered in the response term of paired-associate lists. Distributed practice consisted of 30 sec. between each learning trial and massed practice 4 sec. between each trial. The results showed that with two-letter responses distributed practice facilitated learning when interference in learning the response was high, but not when it was low. Distributed practice did not facilitate the learning of lists in which the response terms consisted of three letters and this was true for all levels of interference.

Experiment III involved lists of 16 pairs of words in which interference was varied within lists by varying the stimulus-response relationships of clusters of words from the same concept categories. Although wide differences in learning occurred as a consequence of differential interference, no facilitation occurred by distributed practice.

The evidence indicates that distributed practice will facilitate acquisition only when interference obtains in response learning, and not when the interference occurs in stimulus-response learning.

References

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. *Cultural norms of verbal items in 43 categories.* Technical Report No. 22, 1957, University of Connecticut, ONR Contract Nonr-631(00).

Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 143–149.

Underwood, B. J. Ten years of massed practice on distributed practice. *Psychol. Rev.*, 1961, **68**, 229–247.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

Underwood, B. J., & Schulz, R. W. Studies of distributed practice: XX. Sources of interference associated with differences in learning and retention. *J. exp. Psychol.*, 1961, **61**, 228–235.

(Received September 2, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 576–582

SEMANTIC SATIATION AMONG BILINGUALS[1]

LEON A. JAKOBOVITS AND WALLACE E. LAMBERT

McGill University

The phenomenon of bilingualism has been studied by linguists long before psychologists considered it of special interest. Weinreich (1953) in an exhaustive review of the literature on this topic reports several studies in which two forms of bilingualism have been identified. Following Osgood's (1952, 1953) development of a theory of meaning which uses a mediation principle as its main concept, Ervin and Osgood (1954) have formulated a psychological theory of bilingualism which incorporates the linguist's notion of two forms of bilingualism, accounting for the differences in types in terms of the manner in which bilinguals acquire and use their two languages. They define a "compound" bilingual as one having two functionally dependent linguistic systems, in contrast to the "coordinate" bilingual who makes use of two functionally independent language systems. The compounds are presumed to have a common mediation process for translated symbols in their two languages, while the coordinate bilinguals are believed to have relatively distinct mediation processes for translation—equivalent symbols in each of their languages. Lambert, Havelka, and Crosby (1958) have provided behavioral evidence in support of this theory.

Lambert and Jakobovits (1960) have recently established a convenient method for measuring a particular type of meaning change known to psychologists as "verbal satiation." This phenomenon can be defined as the decrease in the meaning of a linguistic symbol or sign as a result of its continued presentation. In their experiment, *S*s repeated a particular word continuously for 15 sec. and the amount of change in meaning was determined by comparing the intensity changes of the ratings of the words on various scales of the semantic differential (Osgood, Suci, & Tannenbaum, 1957). For example, the rating of the word FATHER on the 7-point "good-bad" scale would move, after repetition, from Position 1 on the scale ("extremely good") to Position 4 ("neutral or meaningless") when pre- and postrepetition ratings were compared. This phenomenon was viewed as a linguistic analogue of reactive inhibition (Hull, 1943). When a word is repeated the representational mediation process which constitutes the meaning of the word is continuously elicited. Theoretically, reactive inhibition (I_R) develops to the point where it suppresses or at least reduces the intensity of the representational process, resulting in the temporary loss or reduction of the meaning of the word. This reduction in meaning (verbal or semantic satiation) is indexed on the semantic differential as a decrease in the intensity with which the word is rated on the various scales. The size of the reduction is indicated by a "polarity-difference" score which is considered to be an index of "semantic satiability."

Semantic satiation becomes particularly relevant to the functional dependence or independence of the language systems of compound and

[1] This research was supported in part by the Canadian Defense Research Board Grant D77-9401-10.

coordinate bilinguals. When a compound bilingual continuously repeats a word, say HOUSE, the mediation process corresponding to that word should be suppressed, making for a loss of meaning of that word as well as the meaning of its translated equivalent, MAISON, since both words bear the same relation to the suppressed mediators. However, for the coordinate bilingual, translated "equivalents" function relatively independently so that the cross-linguistic satiation effect (hereafter called cross-satiation) should not be exhibited at all or, at least, should be exhibited to a lesser extent than in the contrasting compound case. Here, then, is a potential differentiator between the two types of bilinguals. The experiment which is described below has been specifically designed to test the validity of these theoretical speculations.

METHOD

Materials and Procedure

The method used here to measure semantic satiation was essentially the same as that described in a previous study (Lambert & Jakobovits, 1960). Briefly, this procedure consists of rating words on the semantic differential under two different conditions. In the first, *S*'s normal semantic profile for particular words is determined by having him rate them on a number of semantic (bipolar) scales. Then, in the second or treatment condition, *S* is required to repeat rapidly a word aloud for a period of 15 sec. just before he gives a rating for that word. By comparing the semantic ratings under the two conditions, the amount of decrease in the intensity of ratings attributable to repetition can be determined by means of the polarity-difference score.

The words (see Table 1) and scales were printed on 3 × 5 in. white index cards and placed in a Kardex folder in such a manner as to enable *E* to expose them in a predetermined randomized order. Five semantic scales were used in English for the English words and their French equivalents for the French words. These were: good-bad (bon-mauvais), pleasant-unpleasant (agréable-désagréable), large-small (grand-petit), strong-weak (fort-faible), active-passive (actif- passif). The *S* made his ratings by pointing with a stylus to one of the seven positions along each scale and his responses were recorded. The various treatment conditions which were used are described below and summarized in Table 1.

Normal semantic differential.—The experiment started with the determination of *S*'s semantic ratings under normal conditions for eight words each rated along five scales (making a total of 40 responses in random order). For example, *E* would expose the word FATHER which *S* pronounced aloud once; then *E* exposed a scale (e.g., good-bad) and *S* made his rating. This procedure was followed until all words were rated on all scales. These 40 ratings served as points of comparison for the various treatment conditions described below.

Satiation condition.—Four of the words rated above (two in English and two in French) were presented again (20 responses). The *E* first exposed a word which *S* repeated aloud for 15 sec. at a rate of about two repetitions per sec. Then *E* exposed a scale, a signal to *S* to stop repeating and make his response.

Cross-satiation condition.—Two other words among those rated under the initial normal condition (one in English and one in French) were presented again (making a total of 10 responses). The *E* first exposed a word such as KITCHEN and *S* repeated it aloud for 15 sec. Then *E* exposed a second word which was the translated equivalent word for the first, in this case CUISINE. The *S* pronounced this second word once, after which *E* immediately exposed a scale and *S* made his rating. The *S* had been previously instructed to "always make your response to the second word, not to the first." By comparing the ratings under the present condition with those made under the initial normal condition, the effect of the repetition of its translated equivalent upon the semantic ratings of a particular word (cross-satiation) was determined.

Different-word control condition.—Two other words among those rated under the initial normal condition were presented again. The *E* first exposed a word such as SKY which *S* repeated aloud for 15 sec. Then *E* exposed a second word in the *other* language (FLEUR) and *S* pronounced it once, after which *E* immediately exposed a scale and *S* made his response to the second word. This condition was therefore parallel to the cross-satiation condition except that the word repeated and the word rated did not bear the

TABLE 1

SCHEMATIC REPRESENTATION OF THE PROCEDURE

Condition	Description	Words Used	Examples[a]
Normal semantic differential	Ratings given under normal conditions.	FATHER, COP, VILLE, GARÇON, CUISINE, HOUSE, FLEUR, HATE	"FATHER": "HOUSE":
Satiation	15-sec. repetition of a word followed by rating of that word.	FATHER, COP, GARÇON, VILLE	(FATHER): "FATHER" (GARÇON): "GARÇON"
Cross-satiation	15-sec. repetition of a word followed by rating of its translated equivalent.	CUISINE, HOUSE	(KITCHEN): "CUISINE" (MAISON): "HOUSE"
Different-word control (diff. language)	15-sec. repetition of a word followed by rating of an unrelated word in the other language.	FLEUR, HATE	(SKY): "FLEUR" (BATEAU): "HATE"
Different-word control (same language)	15-sec. repetition of a word followed by rating of an unrelated word in the same language.	MONEY, VÉRITÉ	(SMOKE): "MONEY" (PAYSAN): "VÉRITÉ"

[a] In the examples, the words in parentheses are repeated for 15 sec. and the words in quotation marks are rated.

relationship of translated equivalents, but were semantically unrelated words. This condition was intended to serve as control for the cross-satiation condition and was introduced to determine the effect of the task of repetition per se upon the rating of another word.

It should be mentioned here that the compound group was recalled to the laboratory and tested a second time for reasons to be discussed below (see Discussion). During this second testing a different-word control condition was administered in which the word repeated (SMOKE) was in the *same* language as the word which was rated (MONEY).

The order of presentation of the various treatment conditions was systematically varied to control for and permit identification of possible order and fatigue effects.

Subjects

The *S*s were a heterogeneous group of 62 English-French bilinguals (college students, graduate students, housewives) who were paid for their cooperation. An attempt was made to use "balanced" bilinguals by asking for volunteers who felt "equally proficient in both languages." However, to insure this requirement, Lambert's (1955) automaticity measure was administered to every *S*. This measure involves the use of eight finger keys, the stems of which are differently colored, an exposure apparatus which directs *S* to depress a particular key with directions appearing randomly in either language, and a chronoscope measuring latency of response. A significance test can be applied to test the hypothesis that balanced bilinguals will show a mean differential latency of response which is not significantly different from zero. Two bilinguals who were not balanced were excluded from analysis of results.

In order to determine compoundness or coordinateness among bilinguals, detailed information was obtained from each *S* concerning how, when, and where his languages were acquired. The variables which have been shown by Lambert et al. (1958) to be relevant were used as a basis for classification, i.e., whether *S* had learned the two languages in the same cultural setting or not, whether he has used both languages interchangeably inside and outside the home, etc. On the basis of this interview *S*s were classified as either compound or coordinate, and their results analyzed separately. Separate groups of 30 compounds and 32 coordinates were thus obtained.

Results

The initial polarity scores given in Table 2 were derived by totaling each *S*'s ratings along the 4-point intensity scale and determining a group mean. Thus 16.67 (cross-satiation) is a group mean calculated for 10 ratings per *S*. Since the initial polarity scores could range from 0 to 30, it will be noted that all of the actual scores fall near the median, ranging from 13.94 to 18.07. Consequently it is unlikely that these initial differences influenced the experimental treatment effects to be noted below.

Table 2 also presents the mean change scores which are the differences between initial and final polarity scores. For example the mean change for the compound group after the satiation treatment was −1.27 which indicates a change from an initial polarity score of 16.50 to a final polarity score of 15.23.

Tests of significance are given for departure from zero of each group's mean change, and for differences between mean changes for the compound and coordinate groups on each of the three conditions. (All *t* tests quoted in this experiment are two tailed.) It can be seen that the two groups differ significantly from each other under the cross-satiation condition but not under the other two conditions. This finding substantiates our prediction that compound and coordinate bilinguals will react differently to the cross-satiation treatment. We will now examine the results with respect to each group's behavior under the three different conditions.

Satiation.—The compound group has a negative mean which departs significantly from zero, whereas the coordinate group has a negative mean which fails to reach statistical significance. The result is ambiguous since there is no significant difference between the two groups.

Cross-satiation.—The significant negative mean for the compounds is as expected. The significant *positive* mean for the coordinates was not anticipated. According to the original hypothesis, it was predicted that coordinates will either satiate cross-linguistically to a lesser extent than compounds or will not be affected

TABLE 2

INITIAL SEMANTIC DIFFERENTIAL POLARITY AND MEAN CHANGE IN POLARITY OVER THE SUM OF TWO WORDS AND FIVE SCALES FOR THE COMPOUND AND COORDINATE GROUPS

Condition	Compounds $N = 30$					Coordinates $N = 32$					t^a
	Initial Polarity		Mean Change			Initial Polarity		Mean Change			
	Mean	*SD*	Mean	*SD*	t^b	Mean	*SD*	Mean	*SD*	t^b	
Satiation[c]	16.50	3.90	−1.27	2.74	2.54**	15.03	4.96	−0.25	2.04	0.69	1.64
Cross-satiation	16.67	4.97	−3.27	3.23	5.45***	13.94	5.23	1.16	2.75	2.37**	5.74***
Different-word control	18.07	4.99	−1.30	3.42	2.06**	17.47	4.88	0.53	4.00	0.75	1.89

[a] *t* for average change in polarity between compounds and coordinates.
[b] *t* for departure from zero, two-tailed.
[c] Since there were two English and two French words under this condition, the mean of each pair was used to permit comparisons with the scores in the other conditions.
** $P < .05$.
*** $P < .01$.

by the cross-satiation treatment. It now appears that the coordinate group behaves in exactly the *opposite* manner from the compounds, exhibiting a "generation" of meaning instead of satiation.

Different-word control.—The coordinate group seems to exhibit no systematic effect or change under this control condition, whereas the compounds have a negative mean which departs significantly from zero. However, the difference between the two groups does not reach statistical significance.

A postexperimental interview revealed that none of the *Ss* were aware of the purpose of the experiment. Most of them believed that *E* was investigating "the differences in my reactions in English and French." The *Ss* who stated that their ratings changed (from normal to treatment conditions) thought this was due to the fact that "I forgot what I had done before," or that "I thought of something else the second time." Only 1 *S* stated that his ratings became less extreme under repetition conditions (including control) and attributed this effect to his "confusion." It appears, then, that the semantic satiation effect operates without explicit awareness on the part of *S* of the purposes of the experiment or of the fact that his ratings change in a systematic manner.

Discussion

Although our main prediction has been borne out, namely that compounds cross-satiate, coordinates do not, it is apparent from the pattern of results in Table 2 that our original account of the effect of the three treatment conditions is incomplete and that our theoretical model requires extension. We shall formulate the following two hypotheses which suggest themselves from the present results, but which are also in agreement with some other findings to be mentioned below.

Hypothesis A.—Bilinguals as a group are less susceptible to the satiation effect than are monolinguals. In a pilot study with 27 bilinguals we have used essentially the same method and procedure as that described in the present investigation except that the automaticity measure was not administered and therefore no statements can be made as to their comparative proficiency in the two languages. The combined mean for the 27 bilinguals under the satiation condition was found to be 0.16 ($SD = 2.07$) which is not significantly different from zero ($t = 0.40$). In the present study, when we combine the scores of both compounds and coordinates under the satiation condition ($N = 62$) we obtain a negative mean of -0.76 ($SD = 2.46$) which is also not significantly different from zero ($t = 0.77$). These findings are in striking contrast with experiments on monolingual *Ss* in which a significant satiation effect was repeatedly obtained (Lambert & Jakobovits, 1960).

Added support for this hypothesis comes from unpublished evidence of a slightly different kind. Carroll's (Carroll & Sapon, 1955, Test V) paired-associate test was administered to 18 *Ss* used in another study (see Lambert & Jakobovits, 1960) and their scores on this test were correlated with their polarity-difference scores under the semantic satiation treatment. A product-moment correlation coefficient of -0.66 ($P < .01$) was obtained. This finding suggests that efficient verbal learning (of the paired-associate type) depends on an ability to resist the semantic satiation effect. Bilinguals, and especially balanced bilinguals, have demonstrated their skill in verbal learning and it is likely that they represent a sample of *Ss* who are low on semantic satiability.

Hypothesis B.—An experimental situation in which the bilingual is required to switch from one language to the other is inefficient and inhibitory for the compound, facilitative for the coordinate.

This hypothesis seems to contradict the position taken by Ervin and Osgood

(1954) who state that "theoretically, decoding from a foreign language should be facilitative for the compound system, since different inputs are associated with the same representational reactions or meanings" (p. 143). It seems equally possible, however, that the facilitation provided by the similarities of the mediators is counteracted by the interference effects produced when the two languages are to be used separately, as in ordinary life situations. This is partly recognized by Ervin and Osgood (1954) in their statement that "interference is most likely to occur when the languages are closely related and the cultures and experiences associated with the languages are alike" (p. 141). It seems that in order to reach bilingual fluency, the compound must develop an inhibitory mechanism which would reduce the availability (and therefore the potential interference) of the second language while the first is being used. Then, if he is required to switch from one language to the other, he will behave less efficiently than when he is permitted to function with one or the other of his languages.

The cross-satiation and different-word control conditions used in the present experiment represent a language switching situation. As the compound bilingual repeats a word in one language (e.g., SKY), and is then presented with a word in the other language (FLEUR), the decoding of this second word will be inefficient, reflected as a systematic change toward less intense positions or toward the middle or meaningless point on the semantic scales. The coordinate bilingual, however, will behave more efficiently in the same situation since the two language systems are relatively independent and less subject to interlingual interference. It even appears that the switching process is facilitative, as indicated by the increased polarization of the ratings under language switching situations (i.e., different-word control and cross-satiation). In the case of the compounds, the cross-satiation treatment has a neutralizing effect on the meanings of translated equivalents, whereas for the coordinates the repetitive elicitation of mediators in one language appears to "liberate" the mediators of translated equivalents.

The following deductions can be made from the two hypotheses: (*a*) Since the cross-satiation treatment for the compounds involves both the inhibition of mediators as a result of their repeated elicitation and the inhibition of the decoding process of the second language, the negative mean under this condition should be greater than either the mean of the satiation condition or that of the different-word control in which only one effect is operating at a time. Correlated t tests between these three conditions revealed that the cross-satiation condition has a significantly greater negative mean than that of the satiation condition ($t = 2.25$, $P < .05$) as well as that of the different-word control ($t = 3.28$, $P < .01$), thus supporting the first deduction. Furthermore, it appears that the satiation and different-word control treatments have equal effects since the two means are not significantly different ($t = 0.03$). (*b*) Since language switching is facilitative for the coordinate group, there should be a significant difference between both cross-satiation and different-word control conditions and the satiation treatment (which does not involve language switching). Correlated t tests show that while the difference between the cross-satiation and satiation treatments are significantly different in the predicted direction ($t = 2.14$, $P < .05$), the difference between different-word control and satiation conditions does not reach an acceptable level of significance ($t = 1.18$, $P < .30$). The deduction is thus only partially confirmed. The cross-satiation and different-word control treatments seem to have equal effects ($t = 0.68$). (*c*) If a different-word control condition is administered to the compounds in which the word repeated and the word rated are in the same language, no systematic effect should be observed. As mentioned earlier (see Method) the compound bilinguals were recalled to the laboratory about 5 weeks after the first testing and the new different-word control condition (same

TABLE 3

AVERAGE CHANGE IN POLARITY OVER THE SUM OF TWO WORDS AND FIVE SCALES FOR THE COMPOUND GROUP ON THREE DIFFERENT CONDITIONS

Condition	Compounds (N = 28)[a]			t's Between Conditions	
	Mean	*SD*	t[b]	Same Language	Cross-satiation
Different-word control Different language	−1.29	3.48	1.93*	1.50	2.89***
Different-word control Same language	−0.79	3.14	1.32	—	3.98***
Cross-satiation	−3.14	3.30	4.98***	—	—

[a] Two *Ss* were not available for this part of the testing.
[b] t for departure from zero, two-tailed.
* $P < .10$.
*** $P < .01$.

language) was administered using the same standard procedure as before. The results of this testing as well as those obtained previously with this group are presented in Table 3. It can be seen that both types of different-word control conditions are significantly different from the cross-satiation condition, but they are not significantly different from each other. Although the difference is in the predicted direction (greater negative mean under the different language control), this deduction is not supported to a reliable degree by the present results.

SUMMARY

The phenomenon of semantic satiation was related to the mediation theory of compound and coordinate bilingualism. It was predicted that compound bilinguals, having interdependent language systems, will exhibit a cross-linguistic satiation effect (cross-satiation), while coordinate bilinguals, having functionally independent language systems, would not exhibit the cross-satiation effect. The results supported the prediction that the two types of bilinguals would behave differently and showed that compounds clearly exhibit the cross-satiation effect while coordinates react in the opposite manner. Some additional analyses were discussed which have a further bearing upon the differences in the language structures of compound and coordinate bilinguals.

REFERENCES

CARROLL, J. B., & SAPON, S. M. *Psi-lambda foreign language aptitude battery.* Cambridge, Mass.: Author, 1955.

ERVIN, S. M., & OSGOOD, C. E. Second language learning and bilingualism. In C. E. Osgood & F. Seboek (Eds.), *Psycholinguistics. J. abnorm. soc. Psychol.*, 1954, **49,** Suppl., 139–146.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

LAMBERT, W. E. Measurement of the linguistic dominance of bilinguals. *J. abnorm. soc. Psychol.*, 1955, **50,** 197–200.

LAMBERT. W. E., HAVELKA, J., & CROSBY, C. The influence of language-acquisition contexts on bilingualism. *J. abnorm. soc. Psychol.*, 1958, **56,** 239–244.

LAMBERT, W. E., & JAKOBOVITS, L. A. Verbal satiation and changes in the intensity of meaning. *J. exp. Psychol.*, 1960, **60,** 376–383.

OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49,** 197–237.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

WEINREICH, U. *Languages in contact.* New York: Linguistic Circle of New York, 1953.

(Received September 9, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 583–587

RESPONSE PRETRAINING: DIFFERENTIATION OR AVAILABILITY?[1]

ELI SALTZ

Center for the Study of Cognitive Processes, Wayne State University

Underwood, Runquist, and Schulz (1959) have recently demonstrated that pretraining on the response terms of a paired-associate list facilitated learning of this list. The present paper attempts to examine the mechanisms basic to this effect and to evaluate the relative efficacy of the writer's differentiation theory (Saltz, 1960, 1961) as opposed to the Underwood, Runquist, Schulz position that response pretraining is effective because it increases the availability of the responses.

Underwood, Runquist, and Schulz (1959) have suggested that paired-associate learning is a two-phase process. The first phase involves learning the relevant responses. The second phase involves associating these responses to their appropriate stimuli. Pretraining *S*s on the relevant responses, according to Underwood et al., makes the responses available to *S* and permits the association phase to begin more quickly.

The writer's position is that association is not a critical factor in verbal learning: if *S* had to learn a single pair, learning would occur in very few trials. The critical factor is differentiation of the stimuli and responses to be learned. However, contrary to Gibson's (1940) position, the differentiation focused upon is not between S-R units, but differentiation between stimuli and differentiation between responses. The writer has proposed the following assumptions concerning the differentiation process.

1. Initial differentiation: The initial differentiation of an item, for an *S*, can be defined to be a function of the similarity between this item and other items in a list. The less the similarity, the greater the differentiation.
2. Development of differentiation: Differentiation of the term can be defined as increasing as a function of the frequency with which the term has been presented and responded to (e.g., "familiarization" training, Noble, 1955).
3. Response differentiation: Increased differentiation of the response term, in an S-R pair, is assumed to increase the probability that the term will be emitted as a response—though not necessarily to its appropriate stimulus.
4. Stimulus differentiation: Increased differentiation of the stimulus term is assumed to increase the probability that the stimulus term will elicit its *appropriate* response. In other words, response differentiation is required for emission of a term, but it will not be elicited consistently by its appropriate stimulus until that stimulus is also differentiated.
5. Oscillation of differentiation: Both stimulus and response differentiation oscillate randomly and independently from trial to trial, reducing the effective differentiation from that which would be predicted solely from the initial and developed degrees of differentiation. However, the magnitude of oscillation decreases as a function of increased differentiation.

The specific purpose of the present experiment was to determine if pretraining amounts to anything *more* than making the relevant responses available to *S*.[2] This was done by

[1] The present study is part of a program of research, supported by a grant from the National Science Foundation, concerned with the influence of differentiation on verbal processes.

Alfred Klipple ran *S*s for this study, and Shirley Brown assisted in data analysis.

[2] In fairness to the Underwood et al. (1959) position, it should be pointed out that they do not explicitly deny the possibility that factors other than availability may be operat-

making the responses available on cards to all *S*s during paired-associate learning, and forcing *S*s to use these cards during learning (see Procedure section for method used to ensure use of cards). Half the *S*s had received familiarization pretraining on the responses, half had not. If the effect of pretraining is nothing more than making the responses available to *S*s, then there should be no difference between groups receiving pretraining and groups receiving no pretraining. On the other hand, if pretraining is a predifferentiation device, then pretraining should facilitate learning despite the fact that the responses are available to all *S*s.

It also follows, from the writer's formulation, that one of the major effects of response predifferentiation is to reduce oscillation. This was also tested.

Method

Materials.—The stimuli and responses used in the present study were the same as those used by Underwood, Runquist, and Schulz (1959), but the pairing of stimuli and responses was not identical with theirs since they did not indicate the specific pairings. The criterion learning materials consisted of two lists of 10 pairs each in which the stimuli were Glaze nonsense syllables from 45–53% association value and of low intralist similarity. The responses were two-syllable adjectives. In one list all responses were relatively synonymous; in the other list the responses were unrelated in meaning.

Subjects.—The *S*s were 80 Wayne State University students from the introductory psychology course and were randomly assigned to compose groups of 20 *S*s in each of the four conditions.

Pretraining.—The *S*s in the pretraining conditions were shown the responses they were to learn in the subsequent paired-associate task, using a procedure similar to that of Underwood et al. (1959). The list of responses was presented in five alternate learning and recall trials. During the learning trials, items were presented at a 2-sec. rate and *S* was instructed not to learn the adjectives in order since they would be in a different order on each trial. Five different orders were used. Following each learning trial, *S* was asked to write down as many responses as he could remember, and was given 1 min. for this.

Availability forcing.—During paired-associate learning, a card with all the relevant responses was placed on the top of the memory drum. The responses were numbered serially from 1 at the top of the card to 10 at the bottom. The *S* was instructed to respond by giving the *number* appropriate to the response he felt was correct. Five different orders of responses were used so that the number associated with any given response changed from trial to trial. This obviated the tendency for *S*s to learn a number-response association, and forced *S* to examine the card for each response. The five orders of items on the cards were different from any of the orders used during pretraining and from any of the orders of responses on the memory drum during paired-associate learning.

General procedure.—Half the *S*s received pretraining on the responses to be learned in their criterion paired-associate task. Since Underwood et al. (1959) obtained very similar results with *S*s who received pretraining on irrelevant responses and those who received no pretraining, in the present study *S*s in the no-pretraining conditions were not given pretraining on irrelevant responses. The paired-associate learning was at a 4:4-sec. rate with an 8-sec. intertrial interval. The 4-sec. rate was necessary since preliminary study showed that *S*s could not examine the availability cards and emit a response in 2 sec. Five different random orders of pairs were used, none of which was the same as the order on any pretraining trial or on any of the availability cards. The *S*s received 15 presentations of the paired-associate list, and anticipated the correct responses on the last 14 of these trials.

Results

As can be seen in Fig. 1, results of pretraining in the present study are very similar to those obtained by Underwood et al. (1959), despite the fact that all *S*s in the present study

---

ing in the response pretraining situation. On the other hand, the writer's position is that response differentiation is a more basic variable from which variables such as availability can be predicted.

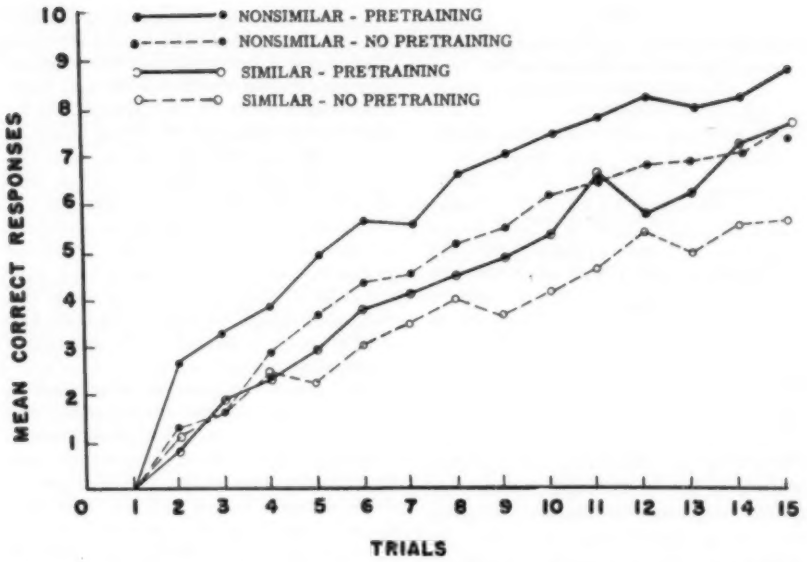

NONSIMILAR - PRETRAINING
NONSIMILAR - NO PRETRAINING
SIMILAR - PRETRAINING
SIMILAR - NO PRETRAINING
MEAN CORRECT RESPONSES
TRIALS
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 2 3 4 5 6 7 8 9 10

FIG. 1. Mean number of correct responses per trial as a function of response pretraining and response similarity.

had the responses available to them on cards during paired-associate learning, and were forced to use these cards in order to respond correctly. As in the Underwood et al. (1959) study, during the first few trials of paired-associate learning, response pretraining was most effective for *S*s who learned material with low intraresponse similarity; by the end of training, pretraining was facilitating both the high and low response similarity groups. A 2 × 2 analysis of variance over the total number of correct responses for the last half of paired-associate training indicated that the effect of pretraining was significant at beyond the .01 level, the effect of response similarity was significant at beyond the .005 level. The interaction between Pretraining and Response Similarity did not approach significance.

A gross oscillation score was calculated for each *S* by summing the number of responses emitted correctly on each trial, which were *not* correctly emitted on the subsequent trial. This gross oscillation score was then divided by the total number of correct responses given by *S* on all trials prior to the trial on which he made no further errors. This yielded an index of the mean amount of oscillation per correct response. Mean oscillation was computed in this manner because some *S*s (particularly in the nonsimilar pretrained group) reached 100% correct responses early in the 15 trials and made no further errors. Dividing by total number of correct responses for such *S*s would give a spuriously low mean oscillation score which would appear to support the writer's hypothesis, but which would actually be merely a reflection of the great number of

TABLE 1

MEAN OSCILLATION PER CORRECT RESPONSE AS AFFECTED BY RESPONSE PRETRAINING AND RESPONSE SIMILARITY

Responses	Pretraining		Nonpretraining		Mean of Means
	Mean	*SD*	Mean	*SD*	
Similar	.205	.125	.307	.201	.256
Nonsimilar	.172	.083	.224	.110	.198
Mean of means	.188		.266		

correct responses emitted during overlearning.

Table 1 shows that response predifferentiation decreases the mean oscillation for both similar and nonsimilar response conditions. In addition, mean oscillation is lower for nonsimilar than for similar material. Both these results are consistent with the writer's formulation concerning the effects of differentiation.

A 2 × 2 analysis of variance (Pretraining × Similarity) indicates that the effect of training on mean oscillation is significant at beyond the .02 level. The effect of similarity falls at approximately the .06 level, and so cannot be considered statistically significant. The interaction yielded an F of less than 1.00.

DISCUSSION

A recent study by Runquist and Freeman (1960) investigated the effects of response pretraining in a "verbal discrimination" situation: two terms were presented simultaneously, and S was to choose one of these. Pretraining on *both* terms in a dyad produced results which were no different from those of a control group with no pretraining. Why did Runquist and Freeman fail to obtain a pretraining effect while such an effect was obtained in the present study? While it is impossible to answer this question with any degree of confidence at this time, it should be pointed out that much less response differentiation was required in the Runquist and Freeman study than in the present study, since in the present study S was required, upon the presentation of a stimulus, to select one response out of a pool of 10 possible responses. The pool of responses was constant from stimulus to stimulus. Runquist and Freeman presented only two responses at a time for S to choose between; the same two responses always appeared together, and neither ever appeared with any other response in a dyad.

The present paper has attempted to interpret the effects of response pretraining in terms of differentiation. Consequently, it is of interest to note that Underwood et al. (1959) reject a differentiation notion because their data do not indicate a greater effect, due to pretraining, for the similar than for the nonsimilar responses. They reason that, if pretraining reduces similarity by means of differentiating the responses, the effect should be greater for similar than nonsimilar material. While this reasoning may be relevant to Gibson's (1940) notion of differentiation, it is not relevant to the formulation presented in the present paper. Within the present meaning of "differentiation," pretraining is effective largely because of the extent to which it reduces oscillation. The relative effect of pretraining on oscillation at various levels of similarity is an empirical question still to be determined.

SUMMARY

The writer has hypothesized that response pretraining is a predifferentiation device. Underwood (Underwood et al., 1959) has hypothesized that it is a procedure for making responses more available during subsequent association learning. The present study investigated the question, does response pretraining facilitate learning even if the responses are made available to S during subsequent association learning. To this end Ss were presented with cards containing all the relevant responses, during the association phase of the experiment. Each response was

numbered, and the numbers were randomly reassigned to responses from trial to trial. The *Ss* were required to emit the number assigned to the appropriate response. Thus all *Ss* were forced to use the availability cards.

The results indicated that pretraining was effective even when the responses were available to all *Ss*. One of the major effects of pretraining proved to be the reduction of mean oscillation per correct response.

References

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

NOBLE, C. E. The effect of familiarization upon serial verbal learning. *J. exp. Psychol.*, 1955, **49**, 333–338.

RUNQUIST, W. N., & FREEMAN, M. Roles of association value and syllable familiarization in verbal discrimination learning. *J. exp. Psychol.*, 1960, **59**, 396–401.

SALTZ, E. Similarity and differentiation in verbal learning. Paper read at Psychonomics Society Meetings, Chicago, 1960.

SALTZ, E. Effect of induced stress on free associations. *J. abnorm. soc. Psychol.*, 1961, **62**, 161–164.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

(Received September 9, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 588–593

THE EFFECT OF STRESS ON STIMULUS DIFFERENTIATION [1]

ELI SALTZ AND WINIFRED J. RIACH

Center for the Study of the Cognitive Processes, Wayne State University

The senior author (Saltz, 1961) has been concerned with the operation of differentiation in human learning. The present experiment is concerned with the effect of stress on such differentiation.

In discrimination learning, degree of differentiation between competing S-R systems is a function of number of reinforcements. Saltz (1961) has postulated that stress acts to break down differentiation between systems. Since the tendency of one system to intrude on another is a function of stimulus similarity (e.g., overlap in common elements between the stimuli to be discriminated), stress should be more disruptive for overlapping systems than for nonoverlapping. This expectation assumes equal differentiation for the overlapping and nonoverlapping systems. Should a greater number of reinforcements be required for differentiation of the overlapping system, the relatively greater disruption of this system resulting from stress would decrease as a function of the greater number of reinforcements. Finally, work by Castaneda and Palermo (1955) suggests that the amount of dedifferentiation produced by stress is greater when original differentiation is low than when it is high.

The following predictions arise from the above formulation. In discrimination learning, when the correct response is dominant, introduction of stress should reduce the probability of correct responses. The greater the similarity between the stimuli to be discriminated, the greater this reduction will be. The reduction will be greatest for similar stimuli when the dominant response is relatively weak.

It should be noted that the predictions arising from the above formulation are different from those made by Castaneda and Palermo (1955), Castaneda (1956), and Palermo (1957), who conceptualize stress in terms of Hullian drive (D). Castaneda and Palermo suggest that stress should facilitate performance when the correct response is highest in the hierarchy, since increased stress produces increased drive, and drive will multiply all existing habit tendencies. On the other hand, if an incorrect response is highest in the hierarchy, stress will have detrimental effects. All three cited studies found that in a simple discrimination task, when the correct response was low in the response hierarchy, performance was hindered by stress more than when the correct response was high in the hierarchy. None of the three studies obtained a significant facilitative effect due to stress. It should be noted, however, that they employed increased pacing (e.g., shorter period of time permitted for response emission) to produce stress. It is not certain that this method of producing stress is equivalent to introducing a painful stimulus or to raising the apprehensiveness of *S*.

The present study was designed to investigate the effects of electric shock as stress in discrimination learning.[2]

METHOD

General design.—The *S*s learned a discrimination task to a low criterion of per-

[1] The present study was supported by a grant to the senior author from the National Science Foundation.

[2] The basic problem in the present study was investigated in a preliminary experiment (Riach, 1958) which showed a decrease in correct responses with shock stress, but was otherwise uninterpretable for reasons of faulty design.

formance, and were then subjected to electric shock simultaneously with the presentation of the discriminanda. A 2 × 2 design was employed: shock-nonshock; similar-dissimilar discriminanda.

Subjects.—The *Ss* were 80 volunteers from the introductory psychology course at Wayne State University. They were randomly assigned to groups of 20 *Ss* (10 men, 10 women) in each of the four experimental conditions.

Apparatus.—The apparatus was modeled after one described by Estes and Burke (1955). Stimuli in the discrimination task were 12 6-v., .25-amp. light bulbs, arranged in a circle with a 9-in. radius on a 2-ft. square display panel painted flat black (see Fig. 1). The patterns of lights were programed on tape. A 7 × 9 in. box was placed 18 in. in front of the display panel, and held two push buttons mounted 5 in. apart on top of the box. The *S*'s task was to learn which of these buttons to depress upon presentation of any stimulus pattern. A 6-v. reinforcement light was located 3 in. above each button; these lights were used to indicate, after each trial, which button should have been pressed. The *E* was informed of the button *S* had depressed by the illumination of a corresponding light on a panel facing *E*. This light could not be activated if *S* responded after illumination of a reinforcement light.

Electric shock was introduced by means of two electrodes attached to *S*'s wrist.

Tasks.—A pattern of three lights on the display panel was shown to *S* for 2 sec. and *S* was to depress either the right or left button on the response box. When the panel lights went off, the reinforcement light indicated the correct button on the response box for 1 sec. There was a 4-sec. interval before the next sequence began. This sequence, which constituted a trial, was electronically controlled.

The highly similar overlap task consisted of making a left-button response when the majority of the three lights appeared at Positions 1–6 on the display panel (see Fig. 1) and a right-button response when the majority of the three lights appeared at Positions 7–12. Stimulus overlap resides in the fact that any stimulus could become connected to both responses, e.g., Stimulus 3 would be part of a pattern indicating the left-button response if one or both of the other stimuli were from Positions 1–6; on the other hand, Stimulus 3 would be part of a pattern indicating the right-button response if the remaining two stimuli were from Positions 7–12.

The nonoverlap task involved pairing any stimulus light with only one response. A left-button response was indicated when three lights at Positions 3–8 were activated; a right-button response was called for by lights at Positions 9–2. Different splits of the lights were used for the overlap and nonoverlap conditions because preliminary work indicated that a vertical split (1–6 vs. 7–12) was too easy for the nonoverlap condition, while a diagonal split (3–8 vs. 9–2) was too difficult for the overlap condition.

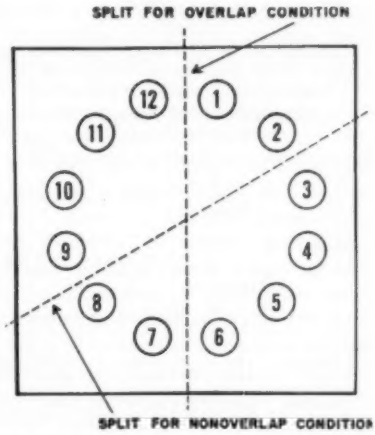


FIG. 1. Illustration of display panel. (Numbers indicating positions of lights, and broken lines indicating splits used for overlap and nonoverlap conditions, are indicated to facilitate understanding of experimental procedure, and were not present on actual display panel.)

Procedure.—The shock device was attached to the nonpreferred hand of each *S*, whether in shock or nonshock condition, before learning was to begin, and shock was adjusted to a point *S* considered to be the pain threshold. The *S* was then told, "now just forget about the armband because I won't apply any stimulation for quite a while." The *S* was also told: "The light will come on over the button you should have pressed for each pattern. This will tell you if you are right or not." All responses were made with the forefinger of the preferred hand. The *S* could rest this hand on the table in front of the response box between responses.

There were 30 different patterns of lights constituting 30 trials for each task before repetition of stimuli began. Responses were tallied in blocks of 12 trials. Nine or more correct responses out of a possible 12 on two

TABLE 1

MEAN BLOCKS OF 12 TRIALS TO CRITERION AND MEAN NUMBER OF CORRECT RESPONSES ON LAST BLOCK BEFORE AND THREE BLOCKS AFTER EXPERIMENTAL TREATMENT

Group	Blocks of Trials to Criterion		Mean Number Correct			
			12 Trials Before		36 Trials After	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Nonoverlap						
Nonstress	3.95	2.21	10.50	.81	32.00	3.54
Stress	3.80	1.54	10.45	.86	30.25	3.69
Overlap						
Nonstress	6.40	3.17	10.20	1.07	29.35	6.95
Stress	7.15	3.66	10.20	1.15	26.75	6.96

separate (though not necessarily consecutive) blocks of trials constituted the criterion for learning. With no pause or instructional indication that a new phase of the testing had begun, *S*s received 36 trials subsequent to reaching the criterion. For an *S* in a stress condition, shock accompanied the first of these 36 trials, plus 11 other randomly selected trials. Shock onset coincided with the onset of the light pattern for a trial, and was removed after ¼–½ sec., before *S* had an opportunity to make a button pushing response.

RESULTS

Table 1 summarizes the number of blocks of trials to reach criterion (including criterional trials) and the performance of the various groups on the 12 trials on which *S*s reached criterion and on the subsequent 36 test trials. As can be seen, the nonoverlap groups reach criterion somewhat earlier than overlap groups. However, there is no significant difference in the strengths of the correct responses for the various groups on the criterion block of 12 trials. Stress produced a relative drop in performance in both overlap and nonoverlap conditions. Also, during the 36 test trials, both overlap groups performed more poorly than the corresponding nonoverlap groups. A difference score was obtained for each *S* between number correct on the criterion trials and on the subsequent 36 trials, and a 2 × 2 analysis of variance was performed over these difference scores. The *F* for stress was significant at beyond the .02 level, the *F* for overlap was significant at beyond the .05 level, and the Stress × Overlap interaction did not approach significance.

Next, the relationship between amount of discrimination learning before the test trials and the performance on the test trials was investigated. It will be recalled that the criterion of learning, before institution of the test series, consisted of *S* making at least 9 correct responses in each of two blocks of 12 trials. A transformed score was calculated for each *S* which was determined primarily by his performance on the last of the two criterion blocks of trials, but which gave some additional weight to performance on the first criterion block. Scores of 9, 10, 11, and 12 correct on the last criterion block of trials were given minimum transformed scores of 1, 5, 9, and 13, respectively. An additional point was added onto the transformed score for every correct response over 9 on the *first* criterion block of trials. Thus, for *S*s who gave 9 correct responses on the last criterion block: 10 correct on the first criterion block was given a transformed score of 2, 11 correct on the first criterion block was given a transformed score of 3, etc. While this transformation was arbitrary, it correlated relatively well with the actual number of correct responses given in the 36 postcriterion test trials. For the nonshock groups, *r* was .40 for the nonoverlap condition, .58 for the overlap condition. For the shock groups, both overlap and nonoverlap conditions produced *r*'s over .70.

For each of the two nonshock groups, a regression was calculated between transformed criterion performance and test performance. For the nonoverlap group, $Y' = .465X + 28.04$ where Y' is predicted test score and X is the transformed criterion score. For the overlap group, $Y' = .71X + 24.88$. Figure 2 shows the relationship between transformed criterion scores and scores on the test trials for the two shock conditions, with the regression line based on the corresponding nonshock conditions indicated. As can be seen, in both overlap and nonoverlap conditions *Ss* with high scores on the preshock criterion trials show almost no drop in performance during the 36 test trials as compared with the regression line for the nonshock group. It is the *Ss* who learned to a low criterion who show the effects of shock. Also, it is apparent that the overlap low criterion *Ss* drop much more compared to the regression line for their nonshock counterparts, than the nonoverlap low criterion *Ss*.

Next the relationship between shock effects and preshock criterion scores were evaluated. The equations, indicated above, predicting test trial behavior from criterion scores for the nonshock control groups, were used to predict test trial behavior for each *S* in the relevant shock groups. These predicted scores were then subtracted from *Ss*' obtained scores. The discrepancy scores obtained in this way indicate the extent to which shock depressed performance for each *S*. Next, all *Ss* in each shock group were divided into two groups, those above the median on the preshocked cri-

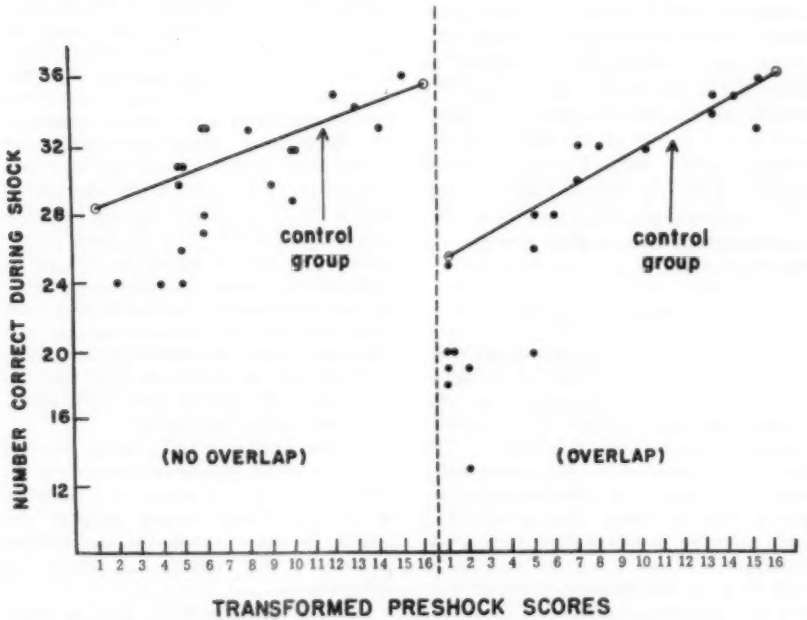


FIG. 2. Scatter plot showing relation between pre- and postshock performance, with regression lines for corresponding nonshock control groups indicated.

terion measure, and those below the median. (It was found that 11 *Ss* in each shock group received a transformed score of 6 or lower—transformed score of 6 corresponds to 10 correct on each of the criterion trials—while 9 *Ss* in each group received a score of 7 or more, and this split came closest to dividing the two groups in half.) For low criterion *Ss*, the mean discrepancy between number correct in the shock condition and number correct predicted on the basis of the nonshock controls was −5.34 in the overlap condition, −2.09 in the nonoverlap condition. For high criterion *Ss*, the mean discrepancy was 0.27 for the overlap condition, −0.59 for the nonoverlap. A 2 × 2 analysis of variance indicates that the shock induced drop in correct responses in the low criterion group is significantly greater than that in the high criterion group at well beyond the .001 level while the interaction between criterion and overlap is significant at beyond the .05 level. Analysis of the simple effects shows that this interaction is a result of the fact that the overlap group with low criterion of original learning dropped in performance proportionately more than the corresponding nonoverlap group as a function of shock.

Discussion

The results of the present study indicate that the introduction of shock as stress produces a decrement in discrimination behavior even if the correct response has become dominant. The decrement appears to be functionally related to degree of differentiation acquired prior to stress. For *Ss* who had high criterion scores before stress was introduced, stress produced no significant drop in performance (compared to relevant nonstress control *Ss*) for either the group with overlapping stimuli or the group with nonoverlapping stimuli. For *Ss* with low criterion scores, however, stress produced a significantly greater decrement in the overlap group than in the nonoverlap group, and both groups showed a significant decrement. These latter results must be treated cautiously for two reasons. First, the differences in criterion scores between high and low criterion *Ss* are based on individual differences in performance rather than controlled variation of degree of learning across *Ss*. Consequently, it is possible that the greater discrepancy in test performance between shock vs. nonshock low criterion *Ss*, relative to shock vs. nonshock high criterion *Ss*, is due to some factor related to learning rate rather than degree of differentiation. Secondly, the overlap and nonoverlap tasks were not equally difficult; as can be seen in Table 1, the overlap conditions required more trials to criterion than the nonoverlap conditions. Since it is assumed that differentiation develops as a function of number of trials, this means that the overlapping S-R systems received more differentiation than the nonoverlapping. From the writers' theoretical position, this suggests that the estimate of greater disruption with shock for overlap than nonoverlap conditions is conservative. Had the amount of reinforcement been equal in the two conditions, relatively more disruption should have occurred in the overlap condition. While the present study could have been designed so that all *Ss* received the same number of pretest trials, a performance criterion was used to assure that the correct response was dominant upon introduction of shock, and yet not so dominant that possible increments in performance with shock could not be demonstrated.

In the present study, there was no evidence for a significant increment in performance in any group as a function of shock. These results support the general position taken by the writers concerning the effects of stress on differentiation.

It should be pointed out that, in terms of the writers' position, the effect of stress on performance can, under appro-

priate circumstances, be facilitative since the effect will depend on the degree of differentiation of the correct S-R system and competing systems. If the correct system is strongly differentiated and all competing systems are very weak, stress should have only a small detrimental effect on the differentiation of the correct system, but should obliterate the weak competing responses; the resultant would be an increase in performance. Increases in noxious stimulation which have facilitated performance in running studies (e.g., Campbell & Kraeling, 1953) and eyelid conditioning studies (e.g., Passey, 1948) can be cited as demonstrating this type of relationship. These studies have of course, in the past been cited as evidence for the multiplicative drive properties of noxious stimulation, i.e., performance = $f(H \times D)$. It is interesting that in this case both the Hullian multiplicative assumption and the writers' assumption that stress breaks down differentiation (divisional rather than a multiplicative assumption) can predict the same phenomena.

Summary

The present study was designed to investigate the effects of stress on previously acquired differentiations. The *S* learned to criterion either a discrimination between stimuli with overlapping stimulus elements, or one between nonoverlapping stimuli. Subsequent introduction of stress in the form of electric shock had the following effects: (*a*) Shock produced a decrement in performance. (*b*) Analysis of individual differences in number of correct responses on the criterion block of trials indicated that the *Ss* who were at high levels of performance prior to the shock showed little decrement. Most of the decrement was contributed by *Ss* who had learned the discrimination to a relatively low (though greater than chance) criterion. (*c*) For *Ss* who were at a high level of performance prior to being shocked, stress produced no differential effect between overlapping stimulus conditions and nonoverlapping. However, there was a large and significant difference for *Ss* who learned to a relatively low criterion, with overlapping groups showing more decrement than nonoverlap. (*d*) No group showed a significant increment as a function of stress. The results were related to theories of differentiation and drive (*D*).

References

Campbell, B. A., & Kraeling, D. Response strength as a function of drive level and amount of drive reduction. *J. exp. Psychol.*, 1953, **45**, 97–101.

Castaneda, A. Effect of stress on complex learning and performance. *J. exp. Psychol.*, 1956, **52**, 9–12.

Castaneda, A., & Palermo, D. S. Psychomotor performance as a function of amount of training and stress. *J. exp. Psychol.*, 1955, **50**, 175–179.

Estes, W. K., & Burke, C. J. Application of a statistical model to simple discrimination learning in human subjects. *J. exp. Psychol.*, 1955, **50**, 81–88.

Palermo, D. S. Proactive interference and facilitation as a function of amount of training and stress. *J. exp. Psychol.*, 1957, **53**, 293–296.

Passey, G. E. The influence of intensity of unconditional stimulus upon acquisition of a conditional response. *J. exp. Psychol.*, 1948, **38**, 420–428.

Riach, W. J. The effect of stress on learned discriminations as a function of stimulus similarity. Unpublished master's thesis, Wayne State University, 1958.

Saltz, E. The effect of induced stress on free associations. *J. abnorm. soc. Psychol.*, 1961, **62**, 161–164.

(Received September 9, 1961)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 594-597

FAMILIARITY AND APPARENT COMPLEXITY OF RANDOM SHAPES[1]

ALVIN G. GOLDSTEIN

University of Missouri

This study attempts to show that judgments of visual complexity are affected by stimulus familiarity. Typically, in previous experiments, perceptual complexity has been defined in terms of the physical characteristics of the stimulus (see Hake, 1957, pp. 59–69 for an excellent review of the many definitions and behavioral measures of figural complexity; also see Attneave, 1955; Elliott, 1958; Fitts, Weinstein, Rappaport, Anderson, & Leonard, 1954; French, 1954; Gregg, 1954; Vanderplas & Garvin, 1959). Complexity is a concept which may also be meaningfully specified by variables associated with *S*. This experiment tested the hypothesis that there is an inverse relationship between the apparent complexity of visual shapes and *S*'s familiarity with these shapes.

METHOD

Stimuli.—Ten random shapes were selected from each of the six categories of stimulus complexity used by Vanderplas and Garvin (1959) in their study of verbal associations to random shapes. The six levels of stimulus complexity were 4, 6, 8, 12, 16, and 24 points.[2] All stimuli were black on white 2 × 2 in. slides. A white cardboard screen 60 in. wide by 40 in. high served as the projection surface. A self-timed, automatic projector was used to present the stimuli. The shapes differed in projected size; the longest axis measured from 12 to 18 in.

Learning trials.—The experiment was performed in two sessions. For Session 1, two shapes from each of the six complexity categories were selected using a table of random numbers. A single capital letter (A to M excluding I) was printed in the lower left corner of these slides. This letter was the "name" of the shape. The *S*s were asked to learn the letter associated with each shape. Trials 1, 2, 4, 6, and 8 were learning trials; *S*s simply tried to learn the letter-shape pairs. Trials 3, 5, 7, 9, and 10 were test trials; the shapes were presented without the letters and each *S* tried to write the correct letter on a special answer sheet. To increase motivation to learn, prizes of $5, $2, and $1 were given to the *S*s with the three highest correct identification scores.

Complexity judgments.—During Session 2, which took place 48 hr. after the learning and test trials, *S*s were shown 36 new random shapes *plus* the 12 shapes (now without identifying letters) they had become familiar with during the previous session. For each of these 48 shapes *S*s were required to judge complexity on a 7-point scale. These instructions were read to the *S*s: "You will be shown black shapes on a white background. I want you to judge how complex each shape is when it is shown to you. Please use the categories listed across the top of your rating sheet."

[1] This research was performed under contract with the International Business Machines Corp. under AF33 (600)31315 with the Aeronautical Systems Center of Air Materiel Command.

[2] There appears to be a discrepancy between what Attneave (1957) calls complexity and what Vanderplas and Garvin (1959) believe Attneave said. Attneave (1957) analyzed the variable, ". . . actual number of turns . . ." (p. 222) whereas Vanderplas and Garvin (1959) say Attneave found that "Complexity refers to the number of points which determine inflections on the perimeter of the shape" (p. 147). Although both investigators seem to be *saying* the same thing, Vanderplas and Garvin do not *mean* exactly what they say. Photographs of all their experimental shapes, arranged in six nominal complexity categories, were presented in the published report. For many shapes the actual number of points is less than the nominal number (e.g., 8-point shapes with five, six, or seven turns). This occurred, of course, as a result of the way in which such random shapes are constructed (Attneave & Arnoult, 1956). Fortunately, this argument (if it is one) is not crucial to the present experiment.

Inquiries were not answered at this time. Following Attneave's (1957) procedure, which was designed to acquaint *Ss* with the range of stimuli to be judged, each of the 48 shapes was shown for 2 sec. before judging began. At this time *Ss*' questions were honored, but answers were only in terms of the instructions. Immediately following this procedure, all the shapes were presented (in a new random order) for 7 sec. each during which time *Ss* made their complexity judgments. The rating scale was the same one used by Attneave (1957): extremely simple, very simple, simple, medium, complex, very complex, and extremely complex.

At the end of Session 2 *Ss* were asked to complete a familiarity scale on each shape. Each shape was now shown for 5 sec., and *Ss* judged whether a shape was (*a*) definitely seen before (that is, before this session), or (*b*) never seen before, or (*c*) uncertain. The purpose of the "uncertain" category was to determine whether shapes other than the learned shapes were thought to be familiar because of similarity between shapes. It must be remembered that simple, randomly constructed shapes often do look very much alike.

An interesting sidelight developed as a consequence of using the familiarity scale. During the second session of a pilot study, identical to the present experiment in procedural details, *Ss* were presented with 48 shapes all different from the 12 they had learned in the first session. When the responses to the familiarity scale were analyzed, it was found that for many of the shapes a large number of *Ss* had checked "definitely seen before." This finding is very similar to Arnoult's (1956) observation that *Ss* work on the principle, "When in doubt, call it familiar" (p. 273).

Subjects.—Two experiments are reported here. The data of the first experiment were obtained from 26 men and women college students. Replication of the experiment was performed on an independent group of 29 men and women students. Each group learned a different set of 12 shapes but both groups were presented with the same 36 unfamiliar shapes. For both groups the sequence of familiar-unfamiliar shapes was identical.

RESULTS

Complexity judgments.—The data of both groups were combined and analyzed together. Therefore there were four familiar and six unfamiliar shapes from each of the six complexity categories. If familiarity has no effect on judgments of complexity, then there should be no statistically significant difference between the mean complexity judgments of the familiar shapes and the unfamiliar shapes.

TABLE 1

MEAN COMPLEXITY JUDGMENTS OF FAMILIAR AND UNFAMILIAR RANDOM SHAPES

No. of Points	Mean Complexity Judgments		
	24 Familiar Shapes	36 Unfamiliar Shapes	All
4	0.75	0.97	0.86
6	1.56	1.71	1.64
8	2.30	2.46	2.38
12	3.10	3.90	3.50
16	3.81	4.30	4.06
24	4.70	5.70	5.20
Mean	2.70	3.17	

Note.—The scale of complexity was as follows: extremely simple, 0; very simple, 1; simple, 2; medium, 3; complex, 4; very complex, 5; extremely complex, 6.

In order to test statistically this proposition, arbitrary values of 0 (extremely simple) to 6 (extremely complex) were assigned to the physical complexity categories. For each shape a mean complexity rating was obtained by tallying all responses according to this system and then dividing by the number of responses (see Attneave, 1957, p. 223 for a discussion of the use of a mean rating). The mean ratings for all the 4-point familiar shapes were then summed and a mean of means was obtained. This procedure was carried out for all the remaining categories of familiar and unfamiliar shapes and the means are presented in Table 1.

The data in the last column of the table clearly support Attneave's (1957) results; psychological complexity is directly related to the number of points used in constructing the random shapes. Four-point shapes were on the average judged to be almost

but not quite "very simple" whereas the 24-point shapes were judged to be a little more complex than "very complex."

In every case the mean complexity judgments for the familiar shapes is less than for the unfamiliar shapes. The mean difference of .47 between familiar and unfamiliar shapes when tested by a t test for correlated scores results in a P between .05 and .02 ($t = 3.189$; 5 df). Separate analysis of the data of the two groups of *Ss* results in essentially identical results.

Familiarity judgments.—The learning trials were specifically designed to allow *Ss* to overlearn the shapes, or at least to learn them well enough to be able to discriminate the learned shapes from the new shapes shown during Session 2. Familiarity ratings were used to determine whether this goal was accomplished. Analysis of these data indicated success; 84% of the combined groups' responses to the 36 unfamiliar shapes were "never seen before," 5% were "uncertain," and 11% were "definitely seen before." Turning to the 24 familiar shapes, 10% of the responses were "never seen before," 2% "uncertain," and 88% were "definitely seen before." Thus, there is evidence that the learning trials did indeed make the shapes familiar to *Ss* even after 48 hr. had elapsed.

Discussion

It may be concluded that judgments of complexity of random shapes, constructed in a particular manner, are inversely related to the familiarity of these shapes. Although one is tempted to generalize this finding to include, for example, more meaningful figural stimuli, there is little experimental evidence to offer in support of such a position. There is, however, experiential evidence which is common to many individuals. For example, upon first looking at a new, unfamiliar machine, piece of apparatus, or engine, it appears to be a complicated mass of rods, levers, and moving parts all unrelated to each other. It is difficult to understand how it operates. Given only a little knowledge of the instrument, gained perhaps by simply "actively" looking at it for a short time, it seems to become simpler, its parts become figures against a background whereas prior to familiarization there was little figure-ground separation.

It is clear that anecdotal evidence is not what is needed to bridge the generalization gap; research with visual stimuli other than randomly generated figures should be initiated (cf. Gibson, 1951). Although the implications of Attneave's (1957) study are clear, it is a little difficult to believe that the number of turns in the contour of a shape is the major—if not the only—determinant of apparent figural complexity. The present experiment, in addition to demonstrating the importance of stimulus familiarity, also emphasizes the need for more critical testing of both *S* and stimulus variables. In designing experiments to do this, consideration should be given to the classical or single variable type of experiment. If a factorial design is employed, as in Attneave's (1957) study, the judging task may be made too simple. In this particular example, the outstanding visible difference between shapes was the number of turns in the contour and it is reasonable to believe that *S*'s judgments were primarily based on this major perceptible difference. It is suggested here that this "visibility" of an experimental variable could be much reduced in a single factor experiment. Obviously, many small experiments, each involving one variable, would have to be conducted.

Vanderplas and Garvin (1959) and Goldstein (1961) have supported the proposition that there is an inverse relation between judged figural complexity and association value. This relationship may be in part a result of an interaction between complexity and familiarity. Consider the proposition that association value and familiarity are directly

related. Thus, a shape should be more familiar if it reminds one of more things or of the same thing more often. If this is assumed to be true, and the present findings suggest that the more familiar shapes appear less complex, then it follows that shapes of high association value should be judged less complex. In other words there is an inverse relation between complexity and association value on a theoretical as well as an experimental basis.

Summary

Judgments of the complexity of 24 familiar and 36 unfamiliar random shapes of matched physical complexity were obtained from 55 *Ss*. Familiarization was accomplished with a 10-trial paired-associate learning session 48 hr. prior to the complexity judging session. The familiar shapes were judged to be significantly less complex than the unfamiliar shapes.

References

Arnoult, M. D. Familiarity and recognition of nonsense shapes. *J. exp. Psychol.*, 1956, **51**, 269–276.

Attneave, F. Symmetry, information, and memory for patterns. *Amer. J. Psychol.*, 1955, **68**, 209–222.

Attneave, F. Physical determinants of the judged complexity of shapes. *J. exp. Psychol.*, 1957, **53**, 221–227.

Attneave, F., & Arnoult, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452–471.

Elliott, L. L. Reliability of judgments of figural complexity. *J. exp. Psychol.*, 1958, **56**, 335–338.

Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, 1954, **51**, 1–11.

French, R. S. Identification of dot patterns from memory as a function of complexity. *J. exp. Psychol.*, 1954, **47**, 22–26.

Gibson, J. J. What is form? *Psychol. Rev.*, 1951, **58**, 403–412.

Goldstein, A. G. Spatial orientation as a factor in eliciting associative responses to random shapes. *Percept. mot. Skills*, 1961, **12**, 15–25.

Gregg, L. W. The effect of stimulus complexity on discriminative responses. *J. exp. Psychol.*, 1954, **48**, 289–297.

Hake, H. W. Contributions of psychology to the study of pattern vision. *USAF WADC tech. Rep.*, 1957, 57–621.

Vanderplas, J. M., & Garvin, E. A. The association value of random shapes. *J. exp. Psychol.*, 1959, **57**, 147–154.

(Received September 12, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 598–604

MEMORY EFFECTS IN CONCEPT LEARNING[1]

EARL B. HUNT

Yale University

In concept learning experiments *S* is shown stimuli selected from a predefined universe. These stimuli (or *instances*) are divided into two or more classes. The task of *S* is to find a rule which can be used to assign any instance in the universe to the appropriate class. When the stimulus universe is defined in terms of possible stimulus variations there is a finite number of possible answers to any concept learning problem (Hovland, 1952). The identification of particular instances as members of specified classes transmits information in the sense that it shows that certain answers cannot be correct. A concept is defined when only one possible answer remains as the definition of each class.

When the instances are presented in series, with only one instance in *S*'s view at any one time, an additional complication is introduced. The information transmitted by the instances must be retained by *S*. From his point of view the concept is not defined until he retains information sufficient to define the concept. This may require presentation of more instances than are necessary to define the concept from *E*'s point of view.

Studies of memory and verbal learning indicate that order of presentation exerts a powerful influence on retention (McGeoch & Irion, 1951). However, there is an important difference between these studies and concept learning. In the usual verbal learning and retention study a particular item (a *key instance*) is inserted in a standard series. After the series is presented there is a test (usually recognition or recall) to determine whether or not *S* can meet some criterion for remembering the key instance. However, in concept learning *S*'s task is to state or use a categorizing rule derived from the information transmitted by the standard series. In order to do this *S* need not remember any particular instance as such. He does have to retain the information transmitted by certain instances.

In order to study information retention, *S*s were presented with a series of instances sufficient to define a concept. The *S*s were then required to identify the class membership of a test series of unlabeled instances, using the categorizing rule they had derived from the first (training) series. The members of the test series were chosen so that their identification depended upon retention of the information transmitted by a key instance in the training series.

EXPERIMENT I

This experiment was designed to provide data revelant to the form of the "information retention curve" as a function of the serial position of the information to be retained.

[1] Based upon a dissertation submitted to Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to express his particular thanks to his advisor, Carl I. Hovland, and to the members of the advisors' and readers' committees. The experiments were performed while the author held General Electric Co. and National Science Foundation fellowships. Research funds were provided by a Ford Foundation grant to Carl I. Hovland for the study of concept learning.

Method

Stimuli.—The stimuli were geometric designs of the type shown in Fig. 1. The particular dimensions and values used to provide stimulus variation (Hovland, 1952) were: type of figure (star, fleur-de-lis, "X", and cross), shape of border (circle or square), color of figure (outline or filled in), and orientation of the "side panels" (0–180°, 45–225°, 90–270°, 135–315°). The side panels are the rectangles extending out from the border of the design.

The designs were dittoed one to an 8½ × 11 in. page. The pages and a cover sheet were stapled into a book. Positive instances of the class to be defined were labeled "Alpha" and negative instances "Bravo."

Information content.—Three sets of instances were used. The "standard set" contained two positive and two negative instances. Three of these were sufficient to eliminate all but two possible concept definitions, "outline figures" or "outline figures and circular borders." The other instance in this set, the key instance, eliminated the definition outline figures. The "redundant instances" were six positive and six negative instances which, taken alone or in combination with the standard set exclusive of the key instance, transmitted sufficient information to restrict the correct definition to either of the two alternatives mentioned. The "test set" consisted of 10 instances whose class membership was not stated. All members of the test set had outline figures and square borders. In order to identify them correctly as negative instances *S* had to retain the information transmitted by the key instance; that instances with outline figures and square borders were not members of the Alpha class.

Table 1 describes all instances used. The instances are listed in order of presentation within a set. The standard set and the test set were always presented as units. The set of redundant instances was split into two or three blocks as required by the experimental design.

TABLE 1

DESCRIPTION OF STIMULI USED, EXP. I

Order in Set	Figure	Panel[a]	Color	Border	Type
		Standard Set			
1	X	135	O	C	P
2	X	135	F	C	N
3	X	O	O	S	N
4	Star	O	O	C	P
		Test Series			
1	Star	90	O	S	?
2	Star	135	O	S	?
3	Fleur	135	O	S	?
4	X	90	O	S	?
5	Cross	45	O	S	?
6	Cross	90	O	S	?
7	Star	45	O	S	?
8	Fleur	45	O	S	?
9	Fleur	O	O	S	?
10	Cross	135	O	S	?
		Redundant Instances			
1	Fleur	O	F	C	N
2	Cross	45	O	C	P
3	Star	90	O	C	P
4	Fleur	135	F	S	N
5	Star	90	F	S	N
6	X	O	O	C	P
7	Cross	90	O	C	P
8	Cross	45	F	S	N
9	Fleur	135	O	C	P
10	Cross	90	F	C	N
11	Star	45	F	C	N
12	Fleur	45	O	C	P

Note.—Abbreviations are: O = outline, F = filled, P = positive, N = negative, C = circle, S = square.
[a] The number in this column is the lowest orientation, in degrees, of the dimension "side panel."

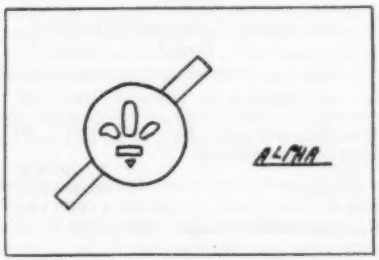


FIG. 1. Typical stimulus: Exp. I.

Design.—The number of instances preceding the key instance and intervening between it and the test series was manipulated by inserting the standard set and the test set into the set of redundant instances. For example, if the standard set is inserted between Redundant Instances 2 and 3 and the test set between Redundant Instances 6 and 7, then four instances are placed in front of the key instance and five intervene between it and the first test instance.

Either 2, 4, 6, or 8 instances preceded the

TABLE 2

MEAN ERROR SCORES: EXP. I

Preceding Instances	Intervening Instances				Mean
	1	3	5	7	
2	2.5	3.7	3.6	5.5	3.8
4	3.5	1.8	3.9	4.4	3.4
6	2.2	5.5	4.9	3.8	4.1
8	2.6	4.7	5.6	3.7	4.2
Mean	2.7	3.9	4.5	4.4	

key instance, and 1, 3, 5, or 7 instances intervened between it and the beginning of the test series. These conditions of the two independent variables were combined in a 4 × 4 factorial design.

Subjects.[2]—The *Ss* were 160 men and women (10 *Ss* per condition) enrolled in the introductory psychology class of Quinnipiac College. The experiment was conducted during regular class periods. Assignment was random within class sections. There were approximately 35 *Ss* in each section.

Procedure.—At the beginning of the class period *E* described a concept learning problem, using numbers as stimuli. The *Ss* were told that they were to solve a conjunctive concept learning problem (Bruner, Goodnow, & Austin, 1956) which involved no more than two dimensions.

Instances were presented by having *Ss* turn the pages of their books on a signal from *E*, which occurred every 30 sec. When *S* encountered a test instance (which was numbered instead of being labeled Alpha or Bravo) he was to indicate, on a separate paper, the class to which the instance belonged.

Redundant instances which were not presented prior to the test series were inserted after it. This made all test books the same length. Since *S* did not know that there would be no further testing after he completed the test series, it was possible, by having all the books the same length, to conduct all experimental conditions simultaneously.

Scoring.—The only conjunctive concept which would correctly categorize all identified instances was "Alpha designs have outline figures and circular borders." If the key instance is excluded, a simpler rule, "All stimuli with outline figures are Alpha," can be derived from the identified instances. It was assumed that in the absence of information to the contrary *S* would utilize the latter, simpler rule. The number of test instances which *S* incorrectly identified as Alpha was used as an operational measure of the difficulty *S* had in retrieving the information transmitted by the key instance. The training series was highly redundant in eliminating possible concept definitions involving other dimensions and values.

This procedure provides a graduated measure of concept learning. It can be argued that such a measure is logically invalid; either *S* has or has not acquired the concept. However, *S*'s (implicit or explicit) statement of a categorizing rule may be only one of several factors in determining his identifications during the test series. Use of a categorizing rule might be masked by "noise" from extraneous variables (e.g., a tendency to match frequency of positive instances in the training and test series). A measure of central tendency would provide maximum sensitivity in determining the information available to *S*. Therefore, the mean number of incorrect identifications was used as the dependent variable.

Results

The mean error scores for the 16 groups are presented in Table 2, Table 3 summarizes an analysis of variance of error scores. Number of errors was a positive linear function of the number of intervening instances. There is no evidence for any effect upon error scores associated with change in the number of preceding instances.

TABLE 3

ANALYSIS OF VARIANCE OF ERRORS: EXP. I

Source	*df*	*MS*	*F*
Preceding instances (P)	3	4.72	.59
Intervening instances (I)	3	26.66	3.36*
Linear	1	60.06	7.69**
P × I	9	13.07	1.65
Residual within groups	144	7.93	

* $P < .05$.
** $P < .01$.

[2] The *Ss* were obtained through the assistance of Robert Evans of Quinnipiac College and Paul A. Lane of the University of Bridgeport.

TABLE 4

MEAN ERROR SCORES: EXP. II

Preceding Instances	Intervening Instances			Mean
	1	9	17	
2	4.1	6.1	7.3	5.8
6	3.2	4.7	4.4	4.1
Mean	3.6	5.4	5.8	

An analysis of types of errors made on the test series did not reveal any striking patterns. In particular, there is no tendency for a particular error to be made in any one experimental condition either more or less frequently than would be expected on the basis of the mean number of errors in that condition.

EXPERIMENT II

In this experiment the number of intervening instances was increased in order to obtain additional empirical knowledge of the form of the information retention curve.

Method

Stimuli.—The stimuli were similar to those used in Exp. I. The dimension "orientation of side panel" was changed to "direction of arrow" and an arrow replaced the "side panel." Triangles, diamonds, and double triangles were added and the fleur-de-lis dropped in the figure dimension. These changes increased the stimulus universe and, hence, the number of redundant instances which could be presented. The stimuli were dittoed and placed on 5½ × 8½ in. pages.

Design.—The same type of design was used as in Exp. I. There were 2 or 6 preceding and 1, 9, or 17 intervening instances.

Subjects.—The *S*s were 54 Bridgeport University introductory psychology students (9 *S*s per condition). One protocol was rejected because of improper assembly of a book. This *S* was assigned the mean error score of his group and the error term degrees of freedom reduced by one.

Results

The mean error scores for each condition are presented in Table 4. The corresponding analysis of variance is summarized in Table 5. The only effect to approach conventional levels of significance ($P < .06$) is the positive linear increase in errors with increase in the number of intervening instances. This is consistent with the results of Exp. I. Paradoxically, the number of errors decreases with an increase in the number of preceding instances. However, this effect is not quite statistically significant, and the difference was in the opposite direction for the same number of preceding instances in Exp. I.

EXPERIMENT III

In Exp. I and II differences in serial position of the key instance were confounded with differences in the length of the training series. Also, differences in the availability of information transmitted by the key instance were ascribed to difficulty in retaining information. Before *S* can retain information, he must acquire it. If the acquisition process varies during learning, serial position effects would appear which were not due to information retention.

In particular, the key instance transmitted information about a particular dimension. Perhaps not all dimensions are scanned on all in-

TABLE 5

ANALYSIS OF VARIANCE OF ERRORS: EXP. II

Source	*df*	*MS*	*F*
Preceding instances (P)	1	39.53	3.51
Intervening instances (I)	2	24.80	2.20
Linear	1	44.60	3.95
P × I	2	4.92	
Residual within groups	47	11.26	

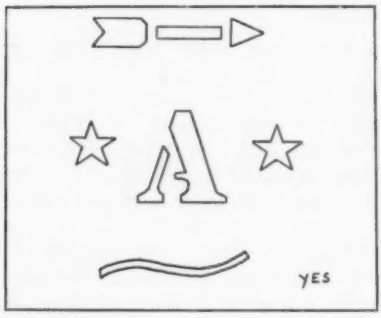


FIG. 2. Typical stimulus: Exp. III.

stances. There is evidence that humans exhibit a systematic order of dominance in scanning dimensions (Dattman & Israel, 1951; Grant, 1951; Heidbreder, 1945). The design of the stimuli and length of exposure interval in Exp. I and II were chosen so that *S*s should have had little trouble in scanning all dimensions. Nevertheless, a more direct test of the effect of transmitting information over different dimensions was desirable.

In Exp. III three key instances, which transmitted information over different dimensions, were presented at various positions in a constant length training series.

Method

Stimuli.—More complex stimuli were used. A typical instance is shown in Fig. 2. There were six dimensions; type of letter in the center (A or C), number of figures on the sides (one or two), type of figure (star, square, or cross), direction of arrow (right or left), number of lines below the letter (one, two, or three), and color of lines (black or white). Training instances were mimeographed on 5 × 8 in. cards, test instances were mimeographed, two to an 8½ × 11 in. page.

Information content.—The training series consisted of 10 positive and 3 negative instances. The negative instances were key instances. Positive instances were marked "Yes" in the lower right hand corner, negative instances were marked "No."

The positive instances all had an arrow pointing to the right, a single figure on either side of the letter, and outline (white) lines below the letter. Thus the positive instances alone restricted the correct concept definition to the inclusive disjunction of these three one-dimensional answers. The three key instances each contained one of the three possible one-dimensional answers, demonstrating that a correct answer involved the union of two or more of the dimensions and values indicated.

The first nine positive instances all had an A in the center. The three negative (key) instances had a C in the center. The tenth positive instance (the thirteenth and last instance of the training series) also had a C in the center. From the results of pretesting, it was felt that *S*s would guess, early in the problem, that "A in the center" was the correct answer. The final instance would be a "positive infirming instance" (Bruner et al., 1956). The *S*s would have to construct a new answer from their memory of previous instances. Comments made by *S*s indicated that the final training instance did force them to develop a new answer. The *S*s were also asked to write a verbal statement of the concept they had derived from the training series. Only 2 *S*s (in different groups) gave answers indicating that they did not realize that the final training instance eliminated the answer "A in the center." These *S*s were replaced.

The test series contained 18 instances. Six of these had an arrow pointing to the right but did not have a single figure beside the letter or outline lines below it. Thus these instances were "positive" by a concept which could be derived from the training series exclusive of one of the three key instances. Similar sets of six test instances were included for the other two key instances.

Design.—Identity and position of key instances (key instances were placed in Serial Positions 1, 5, and 9 of the training series) were combined in a 3 × 3 latin square. In one sequence of training instances Key Instance A appeared in Position 1, B in Position 5, and C in Position 9. This sequence is represented in the first row of the mean error score table (Table 6). The other two sequences of key instances (represented by the second and third rows of Table 6) were CAB, BCA. Each training sequence was given to 10 *S*s, in an example of Lindquist's (1953) Type II design.

Subjects.—Thirty Quinnipiac College students served as paid *S*s. The *S*s were tested in groups of 3 to 8.

Procedure.—Each *S* was given a deck of training stimuli and a test booklet. The same instructions were given as in Exp. I and II except that there was no practice problem. The *S*s from each of the three groups served simultaneously.

Results

The dependent variable was the number of erroneous identifications in the test series which were inconsistent with a particular key instance. The mean number of errors of each type is presented in Table 6, the accompanying analysis of variance in Table 7. The results are not inconsistent with the previously obtained data. Serial position had an effect with a constant length training series. Unfortunately, a stronger statement is not possible because the latin square assumption of noninteracting effects may or may not have been met (Gourlay, 1955). This assumption can be tested by the between-groups and square-uniqueness terms. While neither reaches conventional levels of significance, both are larger than would be desired.

An interaction effect could be produced by selective order in the examination of dimensions. The complex stimuli used should favor some sort of orderly search pattern. This interpretation is consistent with the significant effect of instances, which reflects a reliable difference in the ease with which information is assimilated from different dimensions.

TABLE 6

MEAN ERRORS INCONSISTENT WITH PARTICULAR KEY INSTANCES: EXP. III

Sequence	Position of Instance			Mean
	1	5	9	
ABC	5.0	.5	.2	1.9
CAB	1.8	1.2	.0	1.0
BCA	1.2	2.1	2.2	1.8
Mean	2.7	1.3	.8	

TABLE 7

ANALYSIS OF VARIANCE: EXP. III

Source	*df*	*MS*	*F*
Between *S*s			
Sequence (Groups)	2	7.55	2.46
Within groups	27	3.07	
Within *S*s			
Serial position	2	28.32	5.23*
Instance	2	38.41	7.09*
Sq. uniqueness	2	17.01	3.14
Residual	54	5.42	

* $P < .01$.

DISCUSSION

Considered together, the three experiments present a consistent picture of memory effects during concept learning. The number of errors in identification of instances following presentation of an information-bearing instance increases linearly with the number of instances intervening between the presentation and the test series. The number of instances preceding the key instance does not appear to have a strong effect upon *S*'s ability to retain transmitted information.

These findings may be compared to a related study of memory in concept learning by Cahill and Hovland (1960). They claim evidence for primacy and recency in information retention. Their recency evidence is comparable to "effect of intervening instances." As primacy evidence they show a change in the slope of the function relating probability of use of information to number of intervening instances as the number of preceding instances is varied. This would appear as an interaction effect in the present study. Such an effect was not demonstrated.

Cahill and Hovland (1960) used, as a dependent variable, *S*'s verbal statement of a categorizing rule. In their training series they used only negative instances. In the studies reported here, actual classification be-

havior was observed and both positive and negative training instances were used. To the extent that the two studies have findings in common, their generality is increased. Further research is required to determine the reason for the difference in the findings on the effect (or lack of effect) of preceding instances upon information retention.

Summary

The utilization of information transmitted by a particular instance was investigated as a function of the position of the instance in a series of instances used to define a concept. Utilization of information was determined by observing *S*'s action on a categorization task. An interference effect was found for instances intervening between the information transmitting instance and the beginning of the test.

References

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Cahill, H., & Hovland, C. I. The role of memory in the acquisition of concepts. *J. exp. Psychol.*, 1960, **59**, 137–144.

Dattman, P. E., & Israel, H. E. The order of dominance among conceptual capacities: An experimental test of Heidbreder's hypothesis. *J. Psychol.*, 1951, **31**, 147–160.

Gourlay, N. *F*-test bias for experimental designs of the latin square type. *Psychometrika*, 1955, **20**, 273–288.

Grant, D. A. Perceptual versus analytic responses to the number concept of a Weigl-type card sorting test. *J. exp. Psychol.*, 1951, **41**, 23–29.

Heidbreder, E. Toward a dynamic theory of cognition. *Psychol. Rev.*, 1945, **52**, 1–22.

Hovland, C. I. A "communication analysis" of concept learning. *Psychol. Rev.*, 1952, **59**, 461–472.

Lindquist, E. F. *Design of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1951.

(Received September 12, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 605–611

POSTREST MOTOR LEARNING PERFORMANCE AS A FUNCTION OF DEGREE OF LEARNING[1]

JOHN C. JAHNKE

Miami University

There are extant a number of studies on the effect of amount of prerest practice (degree of learning) on motor learning performance after a rest period (e.g., Adams & Reynolds, 1954; Ammons, 1947; Bell, 1942; Irion, 1949; Kimble, 1949; Wasserman, 1951). However, almost all such studies have involved what might be called massed practice (MP) conditions. The most extreme MP condition in the studies cited above was continuous practice. Less extreme MP conditions involved either 25-5 or 15-5 sec. work-rest cycles; the most highly spaced conditions involved equal amounts of work and rest—either a 60–60 or a 30–30 sec. work-rest cycle. One study (Jahnke, 1958) has explored the effect of degree of learning on postrest performance under practice conditions involving smaller amounts of work than rest with a 15–45 sec. work-rest cycle. Such conditions, involving smaller amounts of work than rest, may be designated by the generic term, distributed practice (DP) conditions. Another DP (15–30 sec. cycle) study by Buxton (1943) also bears on the effect of the degree of learning variable, although the *S*s were practiced to a criterion rather than given a particular amount of practice.

Briefly, the findings of these two studies suggest that the effect of degree of learning under DP conditions is markedly different from its effect under MP conditions. Because of the importance of this suggestion, the two experiments reported below were performed.

METHOD

Experiment I.—Four groups of 20 undergraduate men each were practiced on a pursuit rotor. The naive *S*s were drawn largely from introductory psychology classes. Rate of rotation of the rotor was 60 rpm. Its direction of rotation was clockwise, and it was equipped with a hinged, brass stylus. The brass target was 1 in. in diameter, its surface was flush with the turntable surface, and its center was located 3½ in. from the center of the 11½-in. diameter turntable. Each group received either 8, 17, 22, or 30 prerest trials with the 10–20 sec. work-rest cycle. All *S*s then received 48 hr. rest and 21 additional 10–20 trials. The long rest was introduced to minimize the complicating effects of any residual, undissipated work inhibition, such as Hull's (1943) reactive inhibition (I_R), on postrest performance. The four groups were matched *S* for *S* on the basis of total time on target scores on the first five trials. A running total was kept for each *S* during the matching trials, and *S* was assigned to one of the experimental groups during the 20-sec. rest following the last matching trial. Time on target scores were recorded to the nearest .01 sec. and were read from a Standard Electric clock. Work and rest periods were cycled automatically by a Lafayette Repeat Cycle timer, Model No. 1. The instructions given were those of a previous study (Jahnke & Duncan, 1956).

Experiment II.—Since the results of the lowest degree of learning in Exp. I were the most striking, greater attention was paid in Exp. II to the lower part of the degree of

[1] This research was facilitated by a grant from the Faculty Research Committee of Miami University.

The writer is indebted to S. A. Switzer, who critically read this report. The writer also wishes to acknowledge the assistance of Clifford Hammer, James Korn, and Douglas Sheeran with the collection of data for Exp. I and the assistance of Douglas Sheeran again with the collection of data for Exp. II.

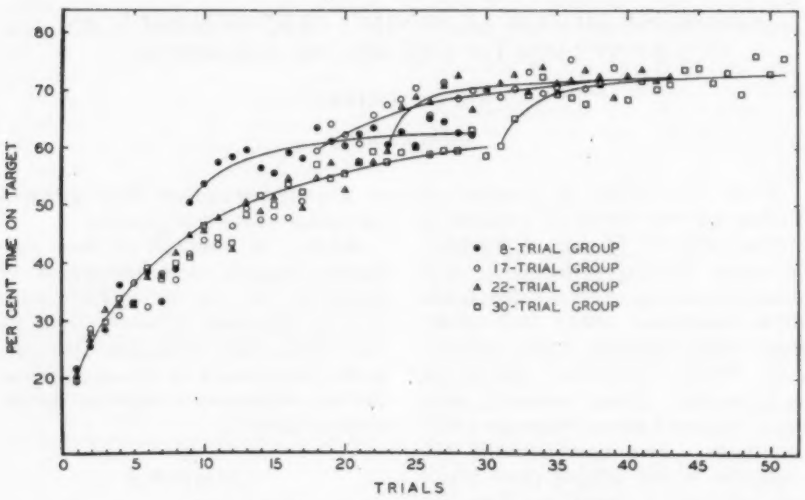


FIG. 1. Performance data for all groups: Exp. I.

learning dimension. Also, the entire dimension was broadened somewhat. Five groups of 20 undergraduate men each were matched S for S on the basis of total time on target during the first four trials. Again a running total was kept for each S during the matching trials, and S was assigned to one of the experimental groups during the 20-sec. rest following the last matching trial. All groups were given a total of 60 trials on the pursuit rotor. Each group received either 4, 8, 16, 24, or 36 practice trials, 48 hr. rest, and its remaining trials. Practice was always with the 10–20 sec. work-rest cycle.

The pursuit rotor was modified somewhat for Exp. II. Its target was $\frac{11}{16}$ in. in diameter and was set flush with the turntable surface. The center of the target was located $3\frac{1}{4}$ in. from the center of the $10\frac{1}{2}$-in. diameter turntable. A spring-loaded stylus with a ball-bearing tip was used, but other details of procedure were the same as for Exp. I.

RESULTS

Experiment I.—The performance data and the smoothed curves fitted to portions of the data are presented in Fig. 1. Inspection of the data shows that the performance curves are quite regular, with a few minor exceptions. The smooth curves fitted to the prerest data of the 30-trial group and the postrest data of all groups are from a single family: a hyperbola with an added constant. (This function was also fitted to the data of Exp. II.) The general equation for this family of curves is:

$$Y = \frac{X - 1}{k + mX} + c,$$

where Y = percentage time on target, X = ordinal number of the practice trial, c = performance on the first trial for which the curve is plotted, and m and k are empirically-determined constants. The curves were fitted by the procedure given by Lewis (1960), and both inspection and tests of goodness of fit suggest the fits are reasonably good.

Inspection of the initial portion of the data of Fig. 1 shows that the matching procedure adequately equated the initial performance of the four groups. A related-measures analysis of variance of performance of these groups over the first eight trials,

during which time practice conditions were identical for all groups, gave $F = 1.06$ for 3 and 57 df ($P > .05$). Terminal performance level before rest was directly related to amount of practice. The index of determination for the fit of the smooth curve to the prerest data of the 30-trial group was .95.

One of the more important results of Exp. I is the function relating amount of reminiscence, or gain in performance over rest, to amount of prerest practice. (In this and the following experiment, reminiscence measures are presented in two ways. The first measure is the difference in performance between the first postrest trial and the last prerest trial for each group. This is a measure of reminiscence uncorrected for practice effects occurring on the first postrest trial. The second measure corrects for practice effects and is the performance difference between the first postrest trial for each group and the trial of the same ordinal number for a control group which has as yet received no long interpolated rest. The averaged performances of all groups which had not yet received rest were used to estimate such a control group's performance. This procedure resulted in a corrected value for each group except the one in each experiment receiving the greatest amount of practice. For these groups, estimates of performance were obtained from their fitted prerest curves.)

Inspection of Fig. 1 and the first portion of Table 1 shows that the groups having smallest amounts of prerest DP exhibit greatest amounts of reminiscence. The correction for practice effects lowered the amounts of reminiscence more for the groups receiving lower degrees of learning than for those receiving higher degrees. All groups of Exp. I show gain in performance over rest when the reminiscence measure uncorrected for practice effects is used; all but the 30-trial group show gain in performance over rest when the alternative reminiscence measure is used. The direct difference t values computed sepa-

TABLE 1

SELECTED PERCENTAGE TIME-ON-TARGET MEASURES FOR ALL GROUPS

Group	Last Prerest Trial		First Postrest Trial		"Estimated" First Postrest Trial[a]	Warm-up[b]
	Mean	*SD*	Mean	*SD*	Mean	Mean
Exp. I						
8-trial	38.95	11.79	50.68	15.30	42.52	9.12
17-trial	51.29	15.02	59.64	12.69	55.85	6.28
22-trial	57.64	15.91	59.34	16.41	57.66	11.62
30-trial	58.80	12.12	60.34	16.03	60.49	9.09
Exp. II						
4-trial	17.92	8.89	27.38	11.07	18.88	8.20
8-trial	22.28	6.96	29.30	12.39	22.17	6.44
16-trial	26.14	8.98	30.46	8.93	33.03	12.70
24-trial	36.23	11.48	34.49	11.73	36.77	13.33
36-trial	43.31	12.87	35.90	9.25	40.46	11.91

[a] Estimates of performance on Postrest Trial 1 if no rest had been given. The difference between this score and the score on Postrest Trial 1 for each group yields a reminiscence measure corrected for practice effects. See text for further details.

[b] These values yield performance on Postrest Trial 5 for each group when added to scores obtained on Postrest Trial 1.

rately for each of the four groups, in order of increasing amounts of practice, for the reminiscence measure uncorrected for practice effects were 5.662, 2.673, .464, and .416, with 19 *df* in each case. The values for the 8- and 17-trial groups are highly significant; the remaining values fail of significance.

Warm-up, defined as improvement in performance over the first five postrest trials, was not found to be systematically related to amount of prerest practice for the conditions of this study. It can be seen in Fig. 1 and again in the warm-up data of the Exp. I section of Table 1 that the initial slope of the curve for the 8-trial group is greater than that for the 17-trial group, lesser than that for the 22-trial group, and about the same as for the 30-trial group. However, when warm-up measures were corrected for practice effects occurring during these five trials, greater ordering was obtained. (Corrections for practice effects for this and the following experiment were estimated from the smoothed prerest curve of the group receiving the greatest amount of prerest practice. The improvement in performance over the appropriate stage of practice for this group was subtracted from the uncorrected warm-up score for each group separately.) The warm-up scores corrected for practice effects were 3.82%, 3.88%, 9.91%, and 7.98% time on target for the groups, given in order of increasing amounts of prerest practice. With the exception of the 30-trial group, whose warm-up measure was somewhat less than that of the 22-trial group, warm-up was directly related to amount of practice. Additionally, it is noted that the groups with the two lowest degrees of prerest learning showed similar amounts of warm-up and that the remaining groups showed larger but also similar, amounts.

Asymptotic performance values were obtained for all four groups. It can be seen in Fig. 1 that the 17-, 22-, and 30-trial groups seem to be converging toward the same value. Asymptotic values for these groups determined from the equations for the fitted curves were 78.44%, 72.75%, and 74.37% time on target, respectively. The comparable computed asymptote of the 8-trial group is much lower (65.11%). The fits of the curves to the postrest data seem acceptable in general, although the fit for the 17-trial group is poor. Indices of determination for these curves were as follows: .80 for the 8-trial group, .29 for the 17-trial group (this low value seems attributable largely to poor performance on postrest Trials 4 and 9), .91 for the 22-trial group, and .82 for the 30-trial group. Additionally, a related-measures analysis of variance was performed on the last postrest trial data. These differences were highly significant ($F = 6.62$, 3/57 *df*, $P < .01$).

Experiment II.—The performance data and smoothed curves fit to portions of the data are shown in Fig. 2. The curve-fitting procedure was identical to that of Exp. I. All curves are again hyperbolic in form. Inspection of the initial portion of the data shows that the matching procedure equated the early performance of the five groups. A related-measures analysis of variance of matching trials performance gave an $F < 1$. Terminal performance level before rest was directly related to amount of practice. The index of determination was found to be .89 for the fit of the smoothed curve to the data of the longest prerest practice group.

The findings bearing on reminiscence in Exp. I were essentially re-

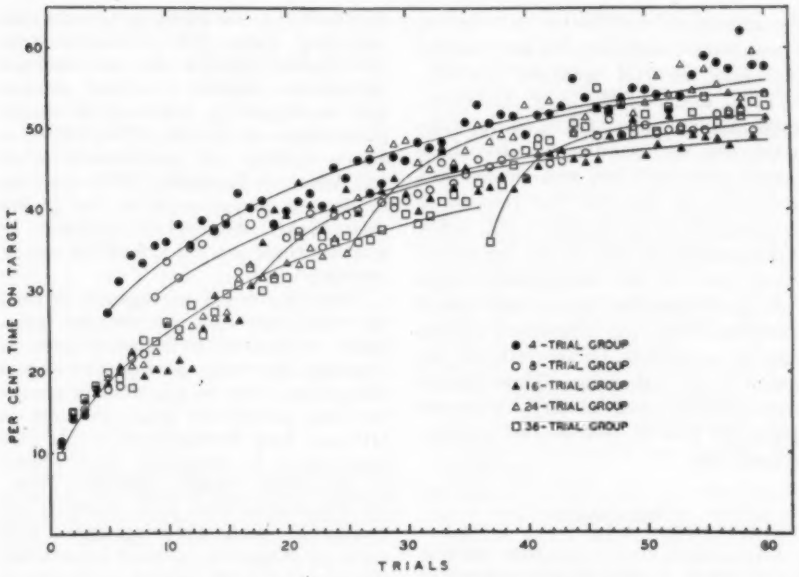


FIG. 2. Performance data for all groups: Exp. II.

peated in Exp. II. Inspection of Fig. 2 and the second portion of Table 1 shows that again groups receiving the smallest amounts of prerest DP showed greatest amounts of reminiscence. In Exp. II, however, losses over rest (forgetting) occurred in the two groups receiving the largest amounts of practice. Direct difference t values were obtained separately for the reminiscence gains (uncorrected for practice effects) for each of the five groups. These values were 4.441, 2.686, 2.227, −.753, and −2.714, with 19 *df* in each case, for the groups in order of increasing practice. All but the fourth value (24-trial group) were significant at or beyond the .05 level.

Warm-up scores, defined as for Exp. I as the gain in performance over the first five postrest trials, are shown in the second portion of Table 1. Warm-up was found to be greatest in the groups with the three highest degrees of prerest learning. As for Exp. I, warm-up values were also corrected for practice effects occurring over the first five postrest trials. The corrected values were 2.46%, 2.18%, 10.08%, 11.56%, and 10.81% time on target for the groups, given in order of increasing amounts of prerest practice. Again the two groups with the lowest degrees of learning showed similar amounts of warm-up and the remaining groups showed larger, but also similar, amounts.

The later postrest practice performances of the five groups do not appear to be systematically ordered. A related-measures analysis of variance of total performance over the last 10 trials showed that performance differences at this stage were not significant ($F=2.06$, 4/76 *df*, $P>.05$). Asymptotic performance values for each group were also determined from

the equations for the fitted curves. These values were similar for the 16-, 24-, and 36-trial groups: 53.34%, 59.12%, and 53.52% time on target, respectively. The indices of determination for the fitted curves for these groups were .87, .89, and .96, respectively. The fits for the remaining groups were less adequate; the index of determination was .27 for the 4-trial group and .62 for the 8-trial group. For this reason, the asymptotic values predicted from the smoothed curves may be unreliable, particularly in the case of the 4-trial group. These values were 70.30% and 65.40% time on target for the 4- and 8-trial groups, respectively.

Discussion

A temporary work inhibition, such as Hull's (1943) I_R, has been conceived to build up rapidly with repeated responding and to dissipate spontaneously with rest. This hypothesis has been quite useful in dealing with important motor learning phenomena, notably the relative depression of performance under MP and, as relates to the present studies, gains in performance occurring over rest. There has been little question of the utility and validity of the concept of I_R; however, results are accumulating which pose difficulty for the concept.

Studies using highly spaced practice conditions apparently do not produce the direct relation between reminiscence and amount of prerest practice demanded by the I_R principle. In both present studies and in an earlier study (Jahnke, 1958), an inverse monotonic relationship between reminiscence and amount of prerest DP was found. Consistent with this finding is that of Buxton (1943), who demonstrated, depending upon response measure, either an inverse monotonic or an inverted U shaped relation between reminiscence and level of mastery. The results of these DP studies do not of themselves negate the I_R principle. They do, however, show that I_R is not the sole factor in reminiscence occurring under DP. Unfortunately, the present studies do not indicate definitively whether a second process, such as forgetting, warm-up, decreased motivation (cf. Kimble, 1949, 1950), or an asymptote of performance effect (cf. Adams & Reynolds, 1954), operated to depress performance in the groups with higher degrees of learning, or whether there is a basic fault in the I_R concept.

Three further points require comment: (*a*) Not only are reminiscence gains likely to be less at higher degrees of learning, but large performance losses (forgetting) may be expected in groups receiving sufficiently large amounts of DP and long interpolated rest. This expectation is congruent with results of an earlier study (Jahnke, 1958). (*b*) The results from Exp. I and II are compatible with the suggestion that warm-up measures corrected for practice effects are directly related to degree of learning. The data corrected for practice effects all suggest that the transition from relatively small to relatively large amounts of warm-up is abrupt, occurring between Trials 17–22 in Exp. I and Trials 8–16 in Exp. II. (*c*) The final point concerns the estimation of asymptotic values from the curves fitted to the postrest trials of the present study. It will be recalled that the asymptotic behavior of the groups of Exp. I suggested that interruption of practice differentially depressed the performance of the group with the lowest degree of prerest learning, while Exp. II showed that this was not necessarily the case. In order to check on the possibility that the small number of postrest trials of Exp. I might have a marked influence on predictions of the asymptotes, the postrest data of Exp. II were fitted as for Exp. I from the first 21 trials. While these curves are not presented here, the results confirm those of Exp. I. It was found that the asymptotes for the groups with the three lowest degrees of learning were very similar (46.33%, 43.63%, and 44.19% time on target for the 4-, 8-, and 16-trial groups,

respectively) and lower than the asymptotes for the two groups with the higher degrees of learning (52.48% and 52.97% time on target for the 24- and 36-trial groups, respectively). Thus, the asymptotic effects noted in Exp. I appear to be attributable to the relatively small number of postrest trials to which the curves were fitted.

While the results of the experiments of this study are not entirely clear on this point, it would seem a reasonable conclusion that groups interrupted at different stages of practice ultimately reach similar postrest performance levels if sufficient postrest practice is given.

Summary

Two experiments were performed to study the effect of degree of learning on postrest performance on the pursuit rotor. The major results were: (*a*) Both prerest and postrest performance data were fitted by a single family of curves. (*b*) Reminiscence was found to be an inverse monotonic function of amount of prerest distributed practice. The implication of this finding for a concept of temporary work inhibition which accumulates with practice is discussed. (*c*) While little or no performance loss over rest (forgetting) was observed in Exp. I, losses were observed in Exp. II. Forgetting occurred in the two groups given large amounts of practice, but was statistically significant only in the group receiving most practice. (*d*) When corrected for practice effects, results were found to be compatible with the hypothesis that warm-up is positively related to amount of prerest practice.

References

Adams, J. A., & Reynolds, B. Effect of shift in distribution of practice conditions following interpolated rest. *J. exp. Psychol.*, 1954, **47**, 32–36.

Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.

Bell, H. M. Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, **54**(1, Whole No. 243).

Buxton, C. E. Level of mastery and reminiscence in pursuit learning. *J. exp. Psychol.*, 1943, **32**, 176–180.

Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

Irion, A. L. Reminiscence in pursuit-rotor learning as a function of rest and of amount of prerest practice. *J. exp. Psychol.*, 1949, **39**, 492–499.

Jahnke, J. C. Retention in motor learning as a function of amount of practice and rest. *J. exp. Psychol.*, 1958, **55**, 270–273.

Jahnke, J. C., & Duncan, C. P. Reminiscence and forgetting in motor learning after extended rest intervals. *J. exp. Psychol.*, 1956, **52**, 273–282.

Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.

Kimble, G. A. Evidence for the role of motivation in determining the amount of reminiscence in pursuit rotor learning. *J. exp. Psychol.*, 1950, **40**, 248–253.

Lewis, D. *Quantitative methods in psychology*. New York: McGraw-Hill, 1960.

Wasserman, H. N. The effect of motivation and amount of prerest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, **42**, 162–172.

(Received September 13, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 612–617

THE CS-UCS INTERVAL IN CONDITIONING SHORT- AND LONG-LATENCY RESPONSES [1]

JOAN E. JONES [2]

University of Sydney

The author has outlined a two-principle theory which predicts the effects of the CS-UCS interval in classical aversive conditioning (Jones, 1962). In brief, this theory is based upon the observation that variation in the CS-UCS interval affects the temporal relationship between (*a*) the CS and the UCR, and (*b*) the CR and the onset of the UCS. With the tentative assumption that UCS-onset is the reinforcing event, the effects of variations in the CS-UCS interval are treated in terms of the joint action of the principles of contiguity (CS-UCR) and of reinforcement (CR-UCS). In both cases a bidirectional gradient is assumed to operate such that the greater the time interval between CS and UCR or CR and UCS the less effective the learning. It is assumed that the efficacy of conditioning at a given CS-UCS interval reflects the additive combination of the overlapping gradients of contiguity and reinforcement.

From this analysis it follows that any variation in experimental procedure which affects either or both of these temporal relationships will affect the function relating the CS-UCS interval to response strength. One such factor is the latency of the response which is being observed. Further, according to the definition given, reinforcement cannot have effect until the CR is evoked (Jones, 1962). Therefore, the contiguity of CS and UCR assumes prime importance in the initial establishing of an overt response. As the CR becomes stronger reinforcement assumes greater importance. This leads to the expectation that the interval function changes during the course of training, so that intervals of different lengths yield learning curves of different shapes.

Although the theory also makes testable predictions about backward conditioning, the present experiment is concerned with forward conditioning and was designed to test the following predictions: (*a*) When groups are trained with a short CS-UCS interval they perform relatively well initially but then show a decrement in performance while groups trained with longer intervals show a less rapid initial rise but perform relatively better later in training. (*b*) In conditioning a short-latency response the acquisition curve shows a relatively rapid initial rise. In the case of a long-latency response a higher asymptote is reached less rapidly. (*c*) Where a response has a relatively long latency and there is some tendency for the CS to elicit that response from the outset a double-peaked function relating the CS-UCS interval to response strength will be obtained. The longer the latency of the response, the greater the separation of the peaks. (*d*) When condi-

[1] Based on a thesis submitted in partial fulfilment of the requirements for the PhD degree, University of Sydney, 1959. This research was supervised by R. A. Champion and part of the apparatus used in this study was provided by Commonwealth of Australia Research Grant No. 2002. Appreciation is expressed to R. B. Bromiley and P. J. Foley for their helpful criticisms of drafts of this article.

[2] Now at the Defence Research Medical Laboratories, Toronto, Canada.

tioning responses which differ in their characteristic latency the interval function for the longer latency response is less sharply peaked and a wider range of intervals is effective.

METHOD

The responses selected were finger movement and the galvanic skin response (GSR). As measured in this experiment, the finger movement had a characteristic latency of 200–250 msec. in response to shock (UCS by instruction) and of 300–350 msec. to a tone (CS). The mean latency of the GSR was 1500 msec. in response to shock (UCS) and 2000 msec. in response to tone (CS). Both responses were conditioned at the same time in each *S* to ensure that they were acquired under identical conditions. Seven CS-UCS intervals were sampled.

Subjects.—The *S*s were 126 volunteers from an introductory course in psychology who were naive regarding the conditioning technique.

Apparatus.—The room containing *E* and the main apparatus adjoined the test room which was lined with a sound reducing material. The test room was well illuminated and circulation of air was maintained by an electric fan. The *S* sat in a comfortable chair equipped with arm and foot rests. The CS was a 2000-cycle, 75-db. tone of 30-msec. duration produced by an electronic oscillator. It was presented to *S* through earphones which at the same time muffled other noises. The UCS was a spike wave ac shock of 24 v., 1.60 ma., and 90 cps of 500-msec. duration, also produced by an electronic oscillator. The shock levels were measured with a 15,000 ohm resistor in place of *S*. The shock was delivered to *S* by means of copper electrodes in the form of spring clamps 15 mm. wide which almost encircled the second and the fourth fingers of the right hand. The duration of CS, UCS, and interonset interval were controlled by electronic timers. On the first and the third fingers of the right hand *S* wore thimble-type electrodes of German silver 30 mm. in length to allow the measurement of the GSR. Electrode jelly was used to ensure good contact at all electrodes. The GSR was measured and amplified by an electronic dc amplifier of the bridge type supplying a constant current of 100 μa. to the electrodes. The four fingers of *S*'s right hand lay on a wooden rest which was depressed by the weight of the fingers. If the fingers were lifted at least 1 mm. a spring caused one contact to close, lighting a lamp on *E*'s control panel, and a second to open, stopping a chronoscope which had started with the presentation of the CS. A four-channel ink-writing recorder, run at 2 mm/sec, recorded the GSR and showed signals corresponding to CS, UCS, and finger movement.

Procedure.—The *S* was seated in the test room and informed of the nature of the stimuli to be presented. The only instruction was to raise the fingers of the right hand from the rest as quickly as possible whenever the shock was felt.

Two pretraining trials were administered to each *S*, the first of which was a UCS (shock) alone, and the second a CS (tone) alone trial. These served to familiarize *S* with the stimuli and showed whether either response was evoked by the CS before training. Twelve training trials followed. These comprised three blocks (a, b, and c) each consisting of three paired presentations of CS and UCS and one test trial (CS alone). The test trial was placed second, third, and fourth in Blocks a, b, and c, respectively, and the order of presentation of the blocks was varied for different *S*s to yield six orders in all. Each order was used for 3 *S*s in each treatment group of 18. The intertrial interval was at least 60 sec. and varied upward of this according to the time taken for *S*'s skin resistance to reach a steady level. The interval was rarely greater than 75 sec. and never exceeded 90 sec. Acquisition curves were obtained showing CR strength measured on Test Trial 1 after 1, 2, or 3 (an average of 2) reinforced trials, on Test Trial 2 after 4, 5, or 6 (average 5), and on Test Trial 3 after 7, 8, or 9 (average 8) reinforced trials. Seven groups were trained, one under each of the following CS-UCS intervals: 20 ± 5, 235, 440, 660, 860, 1045, and 1245 msec.

RESULTS

The method used for recording finger movement yielded only a presence/absence measure so that the index of the strength of the finger CR is the frequency of response evocation. The amplitude of the GSR was recorded as a change in conductance and scores were obtained by expressing these measures as a proportion of *S*'s pretraining response to the CS. This procedure allowed for

TABLE 1

ANALYSES OF VARIANCE OF GSR AMPLITUDES

Source	*df*	*MS*	*F*
CS-UCS interval	6	21.40	.73
Error (between *Ss*)	119	29.40	
Amount of training	2	0.48	.26
Interval × Training	12	3.74	2.03*
Error (within *Ss*)	238	1.84	
Total	377		

* $P < .05$.

individual differences in responsiveness and equated the groups in the pretraining period. A frequency measure was derived by taking scores greater than the median of those obtained from all 126 *Ss* on the three test trials.

An analysis of variance of the GSR amplitude data, summarized in Table 1, yielded a significant interaction between interval and training ($P < .05$ and just greater than .01). Figure 1 shows that the character of this effect is not one of simple divergence of the learning curves but is one of difference in shape. The shorter-interval groups

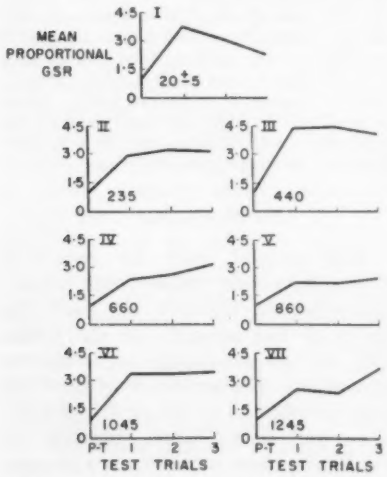


FIG. 1. Performance curves in the acquisition of the conditioned GSR.

reach their maximum performance earlier in training than do the longer-interval groups. Group VI, however, does not show the improvement in Trial 3 which is characteristic of the curves for Groups IV, V, and VII. It is however, unlike the curves for the three shorter-interval groups which characteristically show a decline on later trials. Kendall's tau (τ) was calculated from a ranking of the seven groups in order of mean performance and an arbitrary ranking in order I to

TABLE 2

χ^2 ANALYSIS OF FREQUENCY OF FINGER CRS

Source	χ^2	*df*
CS-UCS interval (I) × Response occurrence (R)	12.21	6
Amount of training (T) × R	0.81	2
Response latency (L) × R	96.39**	1
I × T × R	10.35	12
I × L × R	13.32*	6
T × L × R	16.49**	2
I × T × L × R	7.68	12
Total	157.25**	41

* $P \leq .05$.
** $P < .001$.

VII on each of the three test trials. The obtained values of τ were .33, .24, and $-.41$. The trend in the τ values may be taken as an objective index of the expected form of interaction, the theoretical expectation being that they will change from $+1.00$ to -1.00 over an extended period of training.

The frequency data for finger movement were analyzed by partitioning of χ^2 (Sutcliffe, 1957). The Interval × Training interaction failed to reach the 5% level of significance but was of essentially the same kind

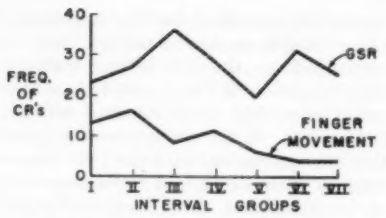


FIG. 2. Frequency of finger CRs as a function of CS-UCS interval.

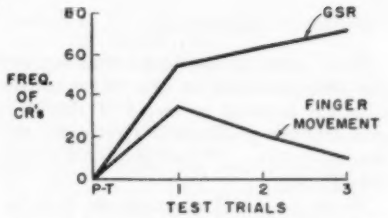


FIG. 3. The acquisition of conditioned GSR and finger movement.

as obtained in the GSR data, τ being .65, .41, and .24 on successive test trials.

Frequency data for the two responses were treated in one analysis which is summarized in Table 2. The Interval × Latency × Occurrence interaction was significant at the 5% level and the Training × Latency × Occurrence interaction at the .1% level. The former interaction, shown graphically in Fig. 2, indicates that the interval effect differs in the case of the two responses. The significance of this interaction may have been enhanced by summing over trials ("nonindependent" measures). However, the same effect appears in the analysis of the number of *Ss* responding (below) which analysis does not suffer the same defect. The Training × Latency × Occurrence interaction (Fig. 3) indicates that the learning curves characteristic of the responses differ in shape. Many more conditioned GSRs were obtained than conditioned finger movements, as indicated by the Latency × Occurrence interaction which was significant at the .1% level.

The frequency data were also compiled in the form of the number of *Ss* responding on one or more test trials during the training period. Table 3 summarizes the analysis and the Interval × Latency × Responders interaction ($.01 < P < .02$) is shown in Fig. 4. Inspection of Fig. 4 indicates that the efficacy of intervals greater than 235 msec. declines markedly for finger movement while no such marked falling off is evident for the GSR. Separate analyses showed the interval effect to be highly significant for finger movement ($\chi^2 = 109.4$, $df = 6$, $P < .001$) but not for GSR ($\chi^2 = 5.2$, $df = 6$, $P \sim .50$).

TABLE 3

χ^2 ANALYSES OF THE NUMBER OF *Ss* RESPONDING

Source	χ^2	*df*
CS-UCS interval (I) × Number of responders (N)	11.25	6
Response latency (L) × N	25.62**	1
I × L × N	16.68*	6
Total	53.55**	13

* $P < .05$.
** $P < .001$.

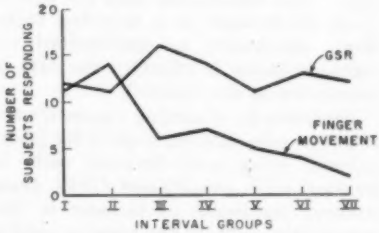


FIG. 4. Number of *Ss* responding as a function of CS-UCS interval.

Discussion

The results of the experiment support the interpretation of the effect of the CS-UCS interval in terms of the joint action of the gradients of contiguity and reinforcement. All four predicted effects were found.

Prediction a.—The amplitude data for the conditioned GSR showed that the shape of the learning curve was affected by the interstimulus interval used in training, short intervals giving a more rapid initial rise but a less well-sustained performance than longer intervals. A similar relationship appeared in the frequency data for finger movement but failed to satisfy the 5% level of confidence. This failure probably results from the insensitivity of the frequency measure. Further confirmation of Prediction a may be found in the literature. The curves presented by Moeller (1954) for GSR and McAllister (1953) for eyeblink show that longer-interval groups approach their performance asymptote at an initially slower rate than shorter-interval groups. Backward conditioning is treated by the theory as a case of true conditioning with a curve similar to that expected with very short forward intervals, i.e., showing a brief initial rise followed by a gradual decline. A curve such as this is described by Spooner and Kellogg (1947). In presenting the results above it was noted that Group VI, which was trained with an interval of 1045 msec., did not conform to neighboring groups conditioned with 660, 860, and 1245 msec. in that it showed no evidence of positive acceleration. This discrepancy does not appear to be attributable to a sampling variation. Apparently an unidentified extraneous factor is affecting the rate of acquisition at this interval.

Prediction b.—Learning curves for the two responses differed, finger CRs (short latency) being most frequent early in training while conditioned GSRs (long latency) continued to increase in frequency throughout the training period (Fig. 3).

Prediction c.—The GSR fulfils the conditions specified for the appearance of a double-peaked function (all *S*s responded to the CS before training) and inspection of Fig. 2 and 4 show that peaks or optima were found at 440 and 1045 msec. If the incongruous acquisition curve obtained with the 1045 msec.-interval should be found to be due to an extraneous factor, the second peak might be disturbed, at least early in training. Similar findings have been reported by Wickens, Gehman, and Sullivan (1959) who conditioned the GSR and obtained an optimum of 500 msec. with a second peak at 1500 msec. Other *E*s (Bierbaum, 1958; Moeller, 1954; White & Schlosberg, 1952) used too few intervals in the critical range to identify a second peak. In the present experiment a slight second peak in the function was obtained with finger movement. This was unanticipated as *S*s included in the study did not give a measurable response to the CS before training. A more sensitive measuring device may have detected a response in some *S*s. This is likely in view of the fact that *S*s were instructed to respond to the shock. The distance between the peaks (235 and 660 msec.) was less with finger movement than with GSR, in accordance with the prediction that the longer the response latency the greater the separation of the peaks when they occur.

Prediction d.—The functions relating CS-UCS interval to measures of CR strength differed as predicted, a more sharply peaked function being obtained for finger movement than for the GSR (Fig. 2 and 4), i.e., a narrower range of intervals was effective in conditioning the response of shorter latency.

The difficulties faced by other theorists in accounting for these effects in terms of the constructs they have proposed, namely, Hull's (1952) stimulus trace, Guthrie's (1935) proprioceptive stimulation, Hilgard and Marquis' (1940) reinforcement principle, and the grouping of stimulus events in cognitive-type theories (e.g., Tolman, 1932) have been discussed elsewhere (Jones, 1962). Briefly, none of these accounts can explain the obtained differences in interval functions

for responses differing in their latency nor do they anticipate the changes in pattern among the curves of acquisition for the various intervals. The basic difference between these theories and the present one is that, while they rely on only one principle, either contiguity or reinforcement, in their account of the CS-UCS interval, this theory explains the effect in terms of the joint action of the two principles.

Summary

An experiment was conducted in which a short- and a long-latency response were conditioned at the same time in seven groups of 18 *Ss*, each group trained with a different CS-UCS interval, ranging from 20 to 1245 msec. The shapes of the learning curves obtained varied with the interval employed and the latency of the response being observed. The variations were consistent with predictions based on the two-principle theory of classical aversive conditioning. Differences between the functions relating performance to the CS-UCS interval for the two responses were also consistent with the theory.

References

Bierbaum, W. B. The temporal gradient in GSR conditioning. *J. gen. Psychol.*, 1958, **59**, 97–103.

Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.

Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Jones, J. E. Contiguity and reinforcement in relation to CS-UCS intervals in conditioning. *Psychol. Rev.*, 1962, in press.

McAllister, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 417–423.

Moeller, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

Spooner, A., & Kellogg, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, **60**, 321–334.

Sutcliffe, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychol. Bull.*, 1957, **54**, 134–137.

Tolman, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.

White, C. T., & Schlosberg, H. Degree of conditioning of the GSR as a function of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

Wickens, D. D., Gehman, R. S., & Sullivan, S. N. The effect of differential onset time on the conditioned response strength to the elements of a stimulus complex. *J. exp. Psychol.*, 1959, **58**, 85–93.

(Received September 17, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 618–625

THE RETENTION OF INDIVIDUAL ITEMS[1]

BENNET B. MURDOCK, JR.

University of Vermont

The present paper reports a study of the short-term retention of individual verbal items. A study by Peterson and Peterson (1959) provided the starting point for the present experiments. In this study a single consonant syllable was presented to *S*, then immediately followed by a three-digit number. The *S* had to repeat the number, count backward by 3's or 4's at a 1-sec. rate for anywhere from 3 to 18 sec., then recall the original syllable. The results showed that the probability of recall decreased exponentially with duration of interpolated activity and approached an asymptote very close to zero.

Three experiments will be reported. Experiment I both replicated the original Peterson and Peterson (1959) study and investigated the effect of varying the nature of the to-be-remembered item. Experiment II was a study of short-term proactive inhibition (PI) which investigated the effect of varying the number of items preceding the to-be-remembered item. Experiment III studied the effect of varying the rate of the interpolated activity.

EXPERIMENT I

Method

This experiment was conducted in four successive sessions spaced about 4 weeks apart. Session 1 was a replication of the Peterson and Peterson (1959) study. The procedure was identical to theirs except for the following conditions: (*a*) the duration of the interpolated activity (i.e., counting backward) was 0, 3, 6, 9, 12, or 18 sec.; thus, we substituted the 0-sec. interval for the 15-sec. interval; (*b*) the metronome was set at rate of 1 beat/sec instead of 2 beats/sec; (*c*) the zero point of the interpolated activity was considered to begin with the first beat of the metronome following the presentation of the nonsense syllable; and (*d*) all *S*s were given sufficient preliminary practice on the interpolated activity to enable them to count backward accurately. As in the Peterson and Peterson (1959) study both latency and accuracy of response were recorded, the consonant syllables were selected from the same pool, and the same controls (counterbalancing of syllables among conditions and conditions among stages of practice) were employed.

Session 2 was an exact duplication of Session 1 except for the to-be-remembered items. Instead of consonant syllables the items were single, monosyllabic, nonhomophonic words selected from the Thorndike-Lorge (1944) list of the 1000 most common words in the English language.

To test whether the large differences obtained between the results of Sessions 1 and 2 could be attributed simply to the number of "chunks" of information (i.e., three unrelated letters in the first session or one common word in the second session; see Miller, 1956) Session 3 was an exact duplication of Sessions 1 and 2 except that the to-be-remembered items were word triads. That is, on each trial *S* was presented three monosyllabic nonhomophonic words from the same pool as above. In selecting the word triads the only stipulation was that the three words could be clearly spoken in the same length of time (i.e., 1 sec.) required to say the three letters of the consonant syllables of Session 1.

As both accuracy and latency of recall were used as dependent variables, Session 4 was conducted to determine whether the duration of the interpolated activity itself would affect response latency. No stimulus item other than the three-digit number was presented; *S* was merely required to say "stop" as quickly as possible after the interpolated activity (counting backward) was terminated.

The same *S*s were used in all four sessions;

[1] This study was supported by a research grant, M-3330, from the National Institutes of Health. Cynthia Marvin conducted many of the experiments and helped in the analysis of the results.

they were students of both sexes from the introductory psychology course who were fulfilling a course requirement. There were 24 *Ss* in each of the first three sessions and 20 *Ss* in the fourth session.

Results

The results of Session 1 clearly confirmed the results of Peterson and Peterson (1959). They presented their results in terms of the proportion of correct responses occurring with latencies less than 2.83 sec. (the overall mean latency for all correct responses). We analyzed our results in the same way and compared them with the values predicted by the formula given in Peterson and Peterson (1959, Fig. 3, p. 195). The differences between the results predicted by this formula and the results actually obtained in Session 1 were small and nonsystematic; the mean (algebraic) difference in proportion was −.025.

The mean proportion recalled at each retention interval is shown in Table 1. These results include all correct responses regardless of latency. To receive credit for a correct response in Sessions 1 and 3, *S* had to recall all three consonants or words in the correct order. Clearly, in all cases forgetting increased with the duration of interpolated activity, and with the longer retention intervals there was very little recall in Sessions 1 and 3.

TABLE 1

MEAN PROPORTION OF CORRECT RECALLS AT EACH RETENTION INTERVAL: EXP. I

Retention Interval	Session 1 CCC Trigrams		Session 2 Words		Session 3 Word Triads	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
0 sec.	.94	.11	.98	.04	.93	.10
3 sec.	.77	.15	.99	.03	.73	.21
6 sec.	.40	.18	.96	.07	.39	.23
9 sec.	.26	.19	.91	.10	.30	.21
12 sec.	.24	.14	.86	.11	.31	.22
18 sec.	.16	.12	.84	.14	.23	.16

As a simple statistical test, for each *S* we determined the total number of items recalled (maximum of 48) over all retention intervals. The mean values were 22.1, 44.3, and 23.0 items with *SD*s of 3.86, 1.99, and 5.98 for Sessions 1 to 3, respectively. The difference between Sessions 1 and 3 was not significant ($t = 0.94$), but the difference between Session 2 and the other two combined was highly significant ($t = 24.9$, $P < .001$).

An analysis of the latency data substantiated the above results. The mean latencies are shown in Table 2 and, in general, latencies increased with the duration of the retention interval. An analysis of variance of the results of Session 4 showed that

TABLE 2

MEAN LATENCY OF CORRECT RECALLS AT EACH RETENTION INTERVAL: EXP. I

Retention Interval	Session 1 CCC Trigrams		Session 2 Words		Session 3 Word Triads		Session 4 Simple RT	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
0 sec.	1.91	0.97	1.13	0.30	1.63	0.28	1.06	0.26
3 sec.	2.01	0.64	1.32	0.31	2.02	0.67	0.85	0.17
6 sec.	2.45	1.01	1.51	0.55	2.32	0.74	0.86	0.18
9 sec.	3.16	1.21	1.61	0.58	2.40	0.72	0.90	0.18
12 sec.	2.49	0.91	1.59	0.62	2.63	1.06	0.92	0.19
18 sec.	3.04	1.18	1.69	0.65	2.79	1.20	0.92	0.25

TABLE 3

MEAN PROPORTION OF CORRECT RECALLS AT EACH RETENTION INTERVAL: EXP. II

Retention Interval	Number of Prior Words					Mean
	0	3	6	9	12	
0 sec.	.995	1.000	.985	.985	.974	.988
3 sec.	1.000	.959	.927	.948	.969	.960
6 sec.	.953	.834	.896	.901	.865	.890
9 sec.	.938	.844	.860	.849	.875	.873
12 sec.	.938	.771	.834	.818	.860	.844
18 sec.	.907	.823	.802	.828	.823	.836
Mean	.955	.872	.884	.888	.894	

the duration of the interpolated activity did have a significant effect ($F = 25.14$, $P < .001$), but most of the difference was due to the 0-sec. retention interval. Thus, the "true" values for the 0-sec. interval for the first three sessions should be slightly lower than those reported. The corrections for differences among the remaining retention intervals are probably too slight to bother with.

To summarize the results of Exp. I: (*a*) they confirm the previous findings of Peterson and Peterson (1959); (*b*) they show that, for three types of items, forgetting, whether measured by accuracy or latency, increased with the duration of the interpolated activity; and (*c*) they show little difference between the retention of three consonants and three words but marked differences between the retention of three "chunks" and one word. The possible confounding of intersession practice effects will be considered with the results of the next experiment.

EXPERIMENT II

Method

This experiment was a study of short-term PI. As in Session 2 of Exp. I the stimulus items were single monosyllabic nonhomophonic words. A given item was preceded by 0, 3, 6, 9, or 12 words, also from the Thorndike-Lorge list of the 1,000 most common words. The preceding words were read at a 1-sec. rate, but always *S* was asked to recall only the stimulus item. The interpolated activity was, as in Exp. I, counting backward for 0, 3, 6, 9, 12, or 18 sec. Each *S* was tested under all conditions. The 5×6 design (five levels of prior words, six retention intervals) required 30 trials (lists) for one replication, and there were eight replications in all. Thus, each *S* was given a total of 240 trials, 60 on each of 4 days. There were 24 *S*s from the same population as before.

In all other respects the procedure was identical with that of Exp. I. There was complete counterbalancing of conditions within *S*s and stimulus material across *S*s. That is, there were eight blocks of 30 trials each so that the experimental conditions appeared equally often at all stages of practice. Within each block the order of the 30 trials was randomized. Each specific stimulus item was assigned equally often to each retention interval and, so far as possible, to each level of number of prior words. The specific words preceding the stimulus item were randomized by starting at a different place on the mimeographed word list for each *S*.

Results

An analysis of intersession practice effects showed that the proportion of words correctly recalled over the 4 days were .87, .87, .92, and .93, respectively. Although the magnitude of the improvement was rela-

TABLE 4

MEAN LATENCY OF CORRECT RECALLS AT EACH RETENTION INTERVAL: EXP. II

Retention Interval	Number of Prior Words					Mean
	0	3	6	9	12	
0 sec.	0.90	1.01	0.98	0.97	1.01	0.97
3 sec.	1.09	1.12	1.17	1.16	1.17	1.14
6 sec.	1.20	1.27	1.33	1.24	1.29	1.27
9 sec.	1.26	1.31	1.35	1.47	1.26	1.33
12 sec.	1.19	1.33	1.35	1.29	1.31	1.29
18 sec.	1.32	1.35	1.37	1.43	1.40	1.37
Mean	1.16	1.23	1.26	1.26	1.24	

tively slight, it was consistent enough to be highly significant ($F = 9.12$, $P < .001$). In Exp. I the proportion of correct recalls for the first session (consonant syllables) was .46, and the proportion of correct recalls for the third session (word triads) was .48. Since the slight improvement was about what would be expected by practice effects alone, the confounding of practice effects and stimulus items probably does not affect the conclusion that consonant syllables and word triads were equally well recalled.

The mean proportion of correct recalls, regardless of latency, at each retention interval in Exp. II is shown in Table 3, and the mean latency of correct recalls is shown in Table 4. Analyses of variance showed that, for each measure, both variables (number of prior words and retention interval) were significant at at least the .01 level. As the retention interval increased, accuracy decreased and latency increased. As the number of prior words increased, accuracy decreased from 0 to 3 words, then increased slightly but consistently from 3 to 12 words. As the number of prior words increased, latency increased from 0 to 3 words but then stayed relatively constant from 3 to 12 words.

When the control condition (0 prior words) is omitted, there were in all 264 extralist intrusions, 161 intralist intrusions, and 81 omissions. Thus, the number of extralist intrusions, intralist intrusions, and omissions occurred in approximately the ratio 3:2:1. An intralist intrusion was considered to be one of the words preceding the to-be-remembered stimulus item, an extralist intrusion was any other word, and an omission was a failure to respond. An analysis of the intralist intrusions showed that almost half (74, or 46%) of these intrusions were the word immediately preceding the to-be-remembered stimulus item, and in general the percentage decreased with increasing remoteness from the stimulus item.

If the control group is included, there were in all 324 extralist intrusions. Of these 324 intrusions 98 (30%) were words that had been correct on one of the preceding trials, 86 were classified as having been mispronounced by *S*, 10 were obvious associations or synonyms of the correct response, and 130 were unclassifiable. Of the 98 intrusions from previous trials 58 had been correct on the immediately preceding trial, 13 on the preceding trial once removed, and the number decreased regularly with increasing remoteness. Thus, in general the analysis of the errors showed marked recency effects for both intralist and extralist intrusions.

To summarize, the results of Exp. II showed that proactive inhibition does occur in the short-term retention of individual items but (at least for accuracy) appears to be a U shaped function of number of previous words.

Experiment III

Method

This experiment was a study of the effect of different rates of interpolated activity on short-term retention of individual items. For this purpose counting backward did not seem to be a suitable type of interpolation; *S*s could not easily count faster than a 1-sec. rate, and a slower rate might allow rehearsal of the original item. Therefore, a new procedure was devised. A list of words of unknown length was read to *S*, and *S* was instructed to remember both the first and the last three words in the list. Requiring *S* to remember the last three words seemed to prevent rehearsal of the first word.

A preliminary experiment was conducted to test the feasibility of this method. There were lists of six different lengths: 4, 7, 11, 16,

21, and 31 words, all nonhomophonic words from the Thorndike-Lorge 1,000 most common. There were 10 counterbalanced replications of the six lengths for a total of 60 trials (or lists) in all. Group testing was used, and all words were read in time with an electric metronome beating at a rate of 1 beat/sec. After each list had been read *Ss* had 20 sec. in which to write down the first and last three words of the list. A "ready" signal preceded by 2 sec. the start of each list. There were 29 *Ss* from the same population as before, all tested in a single session.

An analysis of the results of this preliminary experiment showed that, with one reversal, the proportion of correct recalls of the first word decreased as list length increased. The minimum proportion was .841 for the 31 word list. Statistically, the effect of list length was highly significant ($F = 5.81$, $P < .001$) and, in general, the results were quite comparable to the analogous data on single words in Exp. I and II. Therefore, it was decided to use this method to study the effect of rate of interpolation on short-term retention.

The main experiment was conducted in three sessions and followed closely the procedure of the preliminary experiment. In Session 1 there were six list lengths: 4, 7, 10, 13, 16, and 19 words. The metronome was beating at a rate of 2 beats/sec; starting with the second beat after the first word had been read the remaining words were read either at every beat, every other beat, or every fourth beat. Thus, in effect, the first word was presented for 1 sec. and the remaining words were read at a rate of either 2, 1, or 0.5 words/sec. List length and presentation rate were orthogonal to each other, and there were three replications for a total of 54 trials (or lists). The order of presentation was randomized within each replication.

In Session 2 (actually requiring 2 days to complete) there were three list lengths (7, 13, and 19 words) but five different rates. The metronome was set at a rate of 2.5 beats/sec, and starting with the second beat the remaining words were read at every beat, every other beat, every third beat, every fourth beat, or every fifth beat. Thus, the rates used ranged from 0.5 to 2.5 words/sec. There were eight replications for a total of 120 trials, 60 on each of the 2 days. The order of presentation was randomized within each replication.

In Session 3 there were six list lengths (4, 7, 10, 13, 16, and 19 words) but only two different rates. The metronome was set at a rate of 2.0 beats/sec, and the entire list was read either at a rate of one word every beat or one word every third beat. Thus, the two rates were 2.0 and 0.67 words/sec. These two rates were selected on the basis of the results of Session 2 as those which should yield maximum differences. At the slower rate *Ss* would have more time to rehearse the first word before the rest of the list started; this confounding was deliberately introduced in a further attempt to maximize the differences between the two rates. There were five replications for a total of 60 trials, and the order of presentation was randomized within each replication.

Group testing was used throughout. The same *Ss* were tested in all three sessions (though not in the preliminary experiment). They were 27 students from the same population as before; however, only 21 were tested in Session 3.

Results

In each of the three sessions the proportion of correct recalls decreased as the retention interval increased. Analyses of variance showed that the effect was statistically significant at beyond the .001 level in all three cases.

In Session 1, the proportions of correct recalls were .827, .860, and .829 for the rates of 0.5, 1, and 2 words/sec, respectively. An analysis of variance showed that the differences were not statistically significant ($F < 1.00$). In Session 2, the proportions of correct recalls were .909, .940, .918, .963, and .946 for words read every 1, 2, 3, 4, and 5 beats, respectively. Again, the differences were not statistically significant ($F = 2.10$, $P > .05$). In Session 3, the proportions of correct recalls were almost identical for the two conditions (.894 and .895). Thus, there seems little doubt that, under the conditions of this experiment, the rate of interpolation is not an effective variable governing the short-term retention of individual items.

In Exp. III, *Ss* were asked to recall

not only the first word but also the last three words. Since varying numbers of words preceded the last three words, an analysis of the recall data of the last three words as a function of number of preceding words can give further information about short-term proactive inhibition effects. The proportion of correct recall of the last three words is shown in Table 5. These results show exactly the same pattern as those of Exp. II (see Table 3): an initial drop followed by a gradual rise. The rise from 4 to 16 preceding words was reasonably linear, and the least-squares regression line of the means (bottom row of Table 5) had a slope significantly greater than zero ($t=4.76$, $df=3$, $P<.02$). Thus, these results confirm the conclusion of Exp. II; PI first increases but then decreases as the number of previous words increases.

An analysis of the errors showed omissions of 54%, 63%, and 64%; extralist intrusions of 28%, 25%, and 18%; and intralist intrusions of 18%, 12%, and 18%, each for Sessions 1, 2, and 3, respectively. Averaged across sessions the mean number of omissions, extralist intrusions, and intralist intrusions were 60%, 27%, and 12%, respectively. The corresponding values from Exp. II were 16%, 52%, and 32%, respectively. Thus, in Exp. III there were far more omissions and far fewer intrusions than in Exp. II. In Exp. II the interfering words preceded the to-be-remembered word while in Exp. III the interfering words followed the to-be-remembered word; this difference in procedure may have been responsible for the different ratios of omissions to intrusions found in the two experiments.

TABLE 5

MEAN PROPORTION OF CORRECT RECALLS OF THE LAST THREE WORDS: EXP. III

Session	Number of Preceding Words					
	1	4	7	10	13	16
1	.938	.827	.846	.866	.863	.885
2	—	.874	—	.887	—	.898
3	.916	.867	.890	.895	.892	.905
Mean	.927	.856	.868	.883	.878	.896

TABLE 6

MEAN PROPORTION OF CORRECT RECALLS OF EACH OF THE LAST THREE WORDS: EXP. III

Session	Word		
	Last—2	Last—1	Last
Prelim. exp.	.784	.914	.977
1st Session	.761	.884	.966
2nd Session	.770	.916	.974
3rd Session	.801	.906	.974
Mean	.779	.905	.973

There was a definite serial position effect in the recall of the last three words. Table 6 shows the proportion of correct recalls of the first, second, and third of these last three words (the third word being the last word in the list). The data show the typical serial position effect of free recall (Deese & Kaufman, 1957); there appears to be rather little interexperiment variability.

To summarize, the results of Exp. III showed no effect of rate of interpolation on short-term retention but confirmed the PI findings of Exp. II (and also see Rouse, 1959).

DISCUSSION

The results of the present experiments substantiate and extend the previous word of Peterson and Peterson (1959). With an appropriate interpolated activity, the forgetting of an individual item will increase over a retention inter-

val measured in seconds. Further, the interpolated activity evidently need not be highly similar to the to-be-remembered stimulus item; there would appear to be little formal similarity between the three-digit numbers used in counting backward and the consonant syllables, single words, and word triads used in Exp. I and II.

Given this low similarity, it would seem unlikely that the interpolated activity would interfere with the retention of the stimulus item in the same sense that, for instance, an A-C list interferes with the retention of an A-B list in paired-associate learning. However, the interpolated activity presumably prevents rehearsal of the item, and the results consistently showed that the more time that elapsed without rehearsal the poorer the retention. Thus, the results of the various experiments using the counting backward as an interpolated activity would seem to be more compatible with a decay theory than with an interference theory of forgetting.

On the other hand, the results of Exp. III (using, of course, a quite different interpolated activity) showed that the rate of interpolation was not a significant variable determining forgetting. Given a constant number of interfering items, the total time required to present these items (and hence the time for decay to occur) could vary over a range of three or four to one and still produce the same amount of forgetting. Since in Exp. III it was the number of interpolated items and not their rate of presentation that was the critical variable, these results are more compatible with an interference theory than with a decay theory. Thus, taken as a whole, the present experiments do not provide unambiguous support for either a decay theory or an interference theory of forgetting.

The curvilinear PI function found in Exp. II and III may, at least partly, reflect the "behavioral strategy" (Pollack, Johnson, & Knaff, 1959) adopted by *Ss*. That is, in remembering the last word or words in a list of uncertain length, the more words that have been read the greater the probability that recall of the most recent word or words will be tested; *Ss* can react accordingly. Although such a possibility cannot be completely ruled out it seems unlikely for two reasons: (*a*) As Table 5 shows, the PI function in Exp. III was essentially the same in Session 2 as in Sessions 1 and 3. However, the uncertainty about list length was only half as great in Session 2 and, as Pollack, Johnson, and Knaff (1959) suggest, behavioral strategy would vary with degree of certainty. (*b*) During the course of each session *Ss* would gradually acquire a knowledge of the particular list lengths used, and this knowledge should facilitate the behavioral strategy. However, in Exp. III there was nothing approaching a significant practice effect, either in the preliminary experiment or in Sessions 1-3.

Finally, it is worth noting again the large differences in the retention of consonant syllables and word triads in one case, and individual words in the other case. Over the retention intervals studied, the proportion of correct recalls dropped to about .20 for the former and about .80 for the latter. It can be argued that the consonant syllables and word triads are three items to be remembered while an individual word is only a single item. If this is so, then, the number of items to be remembered would seem to be a significant variable in short-term retention.

Summary

Three experiments were conducted on the short-term retention of individual items. Experiment I confirmed the findings of Peterson and Peterson (1959) and suggested that the number of items or "chunks" of information in the to-be-remembered stimulus item may be a significant variable in short-term retention. Experiment II demonstrated a significant proactive inhibition effect and suggested that the effect may be a U shaped function of number of preceding items. Experiment III showed that the rate of interpolated activity did not have a significant effect on short-term retention. The implications for a decay theory and an interference theory of forgetting were briefly discussed.

References

DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54**, 180–187.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

ROUSE, R. O. Proactive inhibition as a function of degree of practice on the two tasks. *Amer. Psychologist*, 1959, **14**, 385. (Abstract)

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia University, Bureau of Publications, 1944.

(Received September 21, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 626–627

SUPPLEMENTARY REPORT: PERCEPTION OF STIMULI OF VARYING DIMENSIONALITY

PAUL G. SHINKMAN [1]

IBM Research Center, Yorktown Heights, New York

Anderson and Fitts (1958) found that more information was transmitted by numerals superimposed on color patches than by either numerals or color patches alone, when these stimuli were presented for .1 sec. and *S* wrote down what he saw. With numbers alone, and colors alone, the stimuli had from 3 to 8 positions in a row, each containing one symbol from its corresponding alphabet; for the color-numeric stimuli there were from 2 to 4 color patches in a row, with a black numeral in the center of each patch. There were 9 colors, and 9 numerals (0 through 9 exclusive of 1). Generalizing their results, the authors concluded that increasing stimulus dimensionality (in this case combining numbers and colors into a single stimulus complex) improves information transmission.

Method.—The present experiment had three parts: Part I replicated a portion of the original study, and Parts II and III assessed the contribution of differences in visual angle to the main effect in the original work. In Part I three original conditions were repeated; 6 numbers in a row (Cond. N), 6 color patches in a row (Cond. C), and 3 numbers on 3 color patches (Cond. P). These conditions were chosen for replication because they yielded large between-condition differences in the original experiments. In Parts II and III two additional paired configurations of colors and numbers were used in which total visual angle was equated with that of the simple stimuli in Cond. N and C. One contained 6 elements in a row, with 3 color patches on the left and 3 numerals on the right (Cond. AP), and the other simply expanded the 3 superimposed pairs of Cond. P by putting in white spaces (Cond. EP), so that the visual angle for the complete stimulus was the same as for 6 elements in a row (2.68°).

Five *S*s were tested in six sessions each for Part I. In each session 81 stimuli occurred for each of Cond. N, C, and P. The three conditions were counterbalanced over *S*s and sessions. Within conditions, the stimuli were presented in a different random order for each session. In Part II, *S*s served for three sessions, in each of which 81 stimuli occurred for each of Cond. N, C, P, and AP. Part III also involved 5 *S*s for three sessions each, this time under Cond. P and EP.

Other details of method were identical to those of Anderson and Fitts (1958), except as follows: (*a*) In the original study *S*s viewed 2.5 × 3.2 cm. stimulus elements through a shutter device from an average distance of 11 ft. A Harvard tachistoscope was used here and *S*s were 24 in. from elements $\frac{3}{16} \times \frac{1}{4}$ in. in size. Thus, the visual angle subtended by the present stimulus was within 8′ of arc of the original. (*b*) Anderson and Fitts provided 38 ft-c of illumination and a gray wash-out field, whereas the tachistoscope was illuminated by 10 ft-c, and had a pre- and postexposure field of brightness roughly equal to that of the exposure field. (*c*) The present experiment was *S*-paced with *E* recording *S*'s verbal responses; a left-to-right readout order was imposed on *S*. In the original study, *S* wrote down his responses, using whatever order suited him.

Results and discussion.—Table 1 shows information transmitted, by symbol position and condition. The results of Part I are similar to the original Anderson and Fitts data, although the present data show a more striking decrease in transmission from left to right, which may be a function of the imposed readout order. The increase in stimulus dimensionality brought about in the original work by combining colors and numbers involved the introduction of two symbol alphabets into one stimulus, and also a gain in compactness due to the superposition of symbols, making possible a smaller visual angle. Parts II and III of the present study indicate that this smaller angle was in part responsible for the superiority of the complex stimuli over the numbers as reported by Anderson and Fitts.

The present findings also raise two minor questions concerning statements made in the original study. (*a*) The colors were selected by Anderson and Fitts for equal discriminability, but present data on proportions of incorrect responses to the same 9 colors, summed over *S*s, sessions, positions, and conditions, reveal large between-color differences in discriminability. (*b*) Separate informational analysis on each symbol position as

[1] Grateful acknowledgement is made to E. T. Klemmer, under whose direction this research was accomplished.

Now at the University of Michigan.

TABLE 1

AVERAGE INFORMATION GAINED IN BITS

Part	Condition	Symbol Position						Mean/ Symbol
		1	2	3	4	5	6	
I	Numerals	3.03	2.74	2.55	1.45	.90	1.37	2.01
	Colors	2.85	2.66	1.84	.89	.43	.40	1.51
	Pairs							
	Numerals	2.49	1.77	2.09				2.12
	Colors	2.99	2.86	2.65				2.83
II	Numerals	3.09	2.71	2.42	1.68	1.29	1.76	2.16
	Colors	2.98	2.91	2.19	1.29	.79	.67	1.80
	Pairs (Adjacent)							
	Colors/Numerals	2.84	2.60	2.52	1.95	1.44	1.58	2.16
	Pairs (Superimposed)							
	Numerals	2.65	2.26	2.43				2.45
	Colors	3.10	3.02	2.93				3.02
III	Pairs (Superimposed)							
	Numerals	2.88	2.31	2.43				2.54
	Colors	3.09	3.08	3.03				3.07
	Pairs (Expanded)							
	Numerals	1.80	2.37	1.37				1.85
	Colors	3.00	3.06	2.85				2.97

Note.—Averages based on data computed separately for each *S*. Part I: 5 *Ss*, 486 responses per *S* per condition; Part II: 5 *Ss*, 243 responses per *S* per condition; Part III: same as Part II.

done in the original work involves the assumption that errors are independent. For example, transposition errors, which actually transmit information, do not contribute to the obtained estimate of information transmission. On the basis of a pilot study in which transposition errors occurred with an only slightly greater than chance frequency, Anderson and Fitts argued that the corresponding bias in information estimation is negligible. In the present experiment, however, the obtained proportion of transposition errors among all adjacent pairs of errors was .124, compared with a chance expectancy of less than .016. This finding suggests that transposition errors and other positional bias effects are important in a perceptual task of this sort.

In general, the present results show that the Anderson and Fitts data are not only reproducible, but are also stable under minor procedural alterations, including changes in brightness level. Thus the conclusions of the original study are supported.

REFERENCE

ANDERSON, N. S., & FITTS, P. M. Amount of information gained during brief exposures of numerals and colors. *J. exp. Psychol.*, 1958, **56**, 362–369.

(Received September 16, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 627–628

SUPPLEMENTARY REPORT: A SELECTION ARTIFACT IN ROCK'S STUDY OF THE ROLE OF REPETITION[1]

JOANNA P. WILLIAMS

Yale University

Rock (1957) found that *Ss* who had the original items of a list of paired associates presented repeatedly on every trial learned no faster than *Ss* for whom item repetition was omitted, i.e., where new pairs were substituted on every trial for those which *S* had answered incorrectly. He concluded that repetition did not aid learning. There is the possibility, however, that the items in the list were not equal in difficulty and that *Ss* in the nonrepetition group eliminated the difficult items. If so, their final lists would be easier than those learned by the repetition *Ss*. The effect of such a selection factor would oppose that of any possible effect of repetition. The present experiment replicates Rock's study, and in addition determines whether or not such a selection factor does operate in this situation.

[1] These data were collected while the writer was a Teaching Fellow at Harvard University. She is indebted to Fred D. Sheffield for guidance in planning and conducting this research.

TABLE 1
NUMBER OF ERRORS TO CRITERION: Part II

Measure	Group			
	I[a]	II	III	IV
Mean	20.6	21.1	15.2	28.4
SD	9.3	8.6	5.3	11.2

Note.—$N = 20$ in each group.
[a] In Part I, the mean for Group A ($N = 10$) was 19.4, $SD = 8.1$.

Method.—The *Ss* were 100 students at the Harvard Summer School, all volunteers. An additional 14 *Ss* were used, but were eliminated when it became clear that they would not reach criterion within a reasonable amount of time (3 from Groups I, II, and III, 2 from Groups A and IV, and 1 from Group B). The items were 50 randomly combined letter-number pairs. Each letter except *I* appeared as a single, and as a double letter; the numbers were from 1 to 50. Each item was exposed (manually) for 3 sec., with 2 sec. between successive exposures. Training trials, on which both letters and numbers were presented, were alternated with test trials, on which each letter was presented for 5 sec. (with no additional interval between cards) and *S*'s task was to give the appropriate number. No knowledge of results was given during test trials. A 30-sec. interval followed both training and test trials. The criterion was one errorless trial. The procedure was demonstrated by giving *S* one training and one test trial on a list of six word pairs. The *Ss* for whom there was to be substitution of pairs were told to expect such substitution.

In Part I, 20 *Ss* were assigned randomly to one of two conditions. A list of 12 items, chosen at random from the 50-item pool, was presented to each *S*. Group A had Rock's nonrepetition procedure, in which new items were substituted on each trial for those items which *S* had failed to respond to correctly on the previous trial. In Group B, items which were answered *correctly* were eliminated from the list on the next trial, and new ones substituted. (For these *Ss*, the experiment was continued until all 50 items had been presented.) A "final list" was made up from each of these 20 *Ss*. The list from each *S* in Group A was simply that list on which the criterion was reached. The final list from each *S* in Group B was composed of the 12 items on which he had made the greatest number of errors.

In Part II, 80 *Ss* were assigned randomly to one of four groups. A list of 12 items was randomly selected from the item pool for each *S* of Groups I and II. The *Ss* in Group I learned their lists under the same nonrepetition conditions as Group A. Group II learned their lists with repetition of the original items on each trial. These two groups constitute a replication of Rock's experiment. Groups III and IV learned lists under repetition conditions like those of Group II, but their lists were preselected. Each of the first 10 *Ss* in Group III learned one of the final lists from Group A; each of the second 10 learned a final list taken from an *S* of the identically treated Group I. Each final list obtained from Group B was presented to 2 *Ss* in Group IV.

Results.—The mean number of failures to make correct responses before criterion for each group in Part II is presented in Table 1; the overall $F = 7.60$, $df = 3/76$, $P < .001$. No significant difference was found between Groups I and II, thus replicating Rock's original finding. The crucial comparison is between Groups II and III: Group III made significantly fewer errors than did Group II ($P < .05$), indicating that the lists learned by Rock's nonrepetition group may well have been easier than those learned by his repetition group. Also, Group IV made a significantly greater number of errors than did Group II ($P < .01$), showing that items could be selected which were more difficult to learn than a random list.

One important source of selection was noted: items which included a low number (1–15) were more likely to be found in the lists on which *Ss* reached criterion than items with numbers 16–50 ($\chi^2 = 7.77$, $df = 1$, $P < .01$). Items composed of a low number were less likely to appear in those lists found to be more difficult to learn than random lists ($\chi^2 = 14.33$, $df = 1$, $P < .01$).

These results suggest that there may indeed be an important selection bias in the Rock experiment, which would be expected to enhance the performance of the nonrepetition group and to balance any possible effects of the repetition received by the other group. Thus Rock's results cannot be interpreted as evidence for all-or-none learning. There is probably a still greater amount of selection in the Rock procedure than is demonstrated here, because of the likelihood of selection on the basis of idiosyncratic differences in difficulty.

REFERENCES

ROCK, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, 70, 186–193.

(Received October 31, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 629

SUPPLEMENTARY REPORT: ALTERNATION IN A T MAZE WITH HABIT HELD CONSTANT[1]

MERRELL E. THOMPSON AND EDWIN J. MARTIN

New Mexico State University

In an earlier study (Thompson, 1960) designed to check the role of stimulus, reinforcement, and choice conditions in alternation behavior, and to explore the possible interaction of these variables, some tentative conclusions were drawn concerning the importance of reinforcement and the interaction of the stimulus variable with conditions of reinforcement. Although indirect verification was also found for a number of deductions from the senior author's (Thompson, 1960), two-factor theory of inhibition, certain factors in the design made for difficulty in rigorous interpretation of the results. For example, an attempt to check Glanzer's (1953) study unfortunately contaminated the data by use of both constant and reversed stimuli. The influence of the spaced trials upon massed trial behavior was indeterminable, and habit to the various alternatives was uncontrolled (choice of the black or white alleys and right or left turns ranged from 2% to 98% for *Ss* in the free choice groups and from 27% to 73% for the *Ss* in the forced choice groups by the last day of the experiment).

Method.—The present study was designed to give a simple, more concise test of three of the deductions from the theory, as well as to check upon the earlier conclusions regarding the role of the reinforcement and stimulus variables and their interaction. It differed from the earlier study in the following respects: (*a*) all trials were given from opposite start boxes, thus presenting only reversed stimuli; (*b*) only massed trials were employed; (*c*) the first two trials each day (8 days) were free for all *Ss*; and (*d*) habit to the various alternatives was kept constant by administering two forced trials, following the two daily free trials, in one of the following sequences: RRLL, LLRR, RLLR, and LRRL (expressed in terms of the stimulus component: WWBB, BBWW, WBBW, and BWWB).

The present study replicates the original in so far as (*a*) the two independent variables (stimulus conditions and reinforcement conditions) were manipulated at the same values; (*b*) the dependent variable of alternation was repeated; (*c*) Deductions 3, 4, and 5 from the two-factor theory were under test (Deductions 1, 2, and 6, tested in the earlier study, are undergoing further study in two additional investigations); and (*d*) the apparatus and procedure remained unchanged, except as described above. The *Ss* were 80 rats of the Holtzman strain, 70–90 days old.

TABLE 1

MEAN NUMBER OF RESPONSE ALTERNATIONS

Group	Conditions	Mean	t[a]
I	B-B +	3.15	2.18*
II	B-B ○	5.35	4.09**
III	B-W +	3.35	1.81
IV	B-W ○	3.30	2.19*

[a] Tests were made comparing mean number of alternations with chance expection.
* $P < .05$.
** $P < .001$.

Results.—Table 1, which shows the mean number of response alternations for the four groups over the 8 days, reveals that Group II alternated responses significantly above chance. Groups I and IV showed response alternation significantly below chance, i.e., they alternated stimuli significantly above chance.

A three-dimensional analysis of variance, using the four replications as the third variable was performed after Bartlett's test had shown no significant heterogeneity of variance. As in the earlier study, the main effect of reinforcement condition was significant ($F = 10.27$, $df = 1/3$, $P < .05$). Again also stimulus condition was not significant as a main effect, but it interacted with conditions of reinforcement ($F = 32.46$, $df = 1/3$, $P < .025$). In t tests between various groups, significant differences occurred only when Group II was compared with some other group.

As in the 1960 study, Deductions 4 (that nonrewarded *Ss* will alternate stimuli and repeat responses in a Cross T maze when the maze arms are dissimilar) and 5 (that nonrewarded *Ss* will alternate responses and repeat stimuli in a Cross T maze when the arms are similar) were verified. Support was not found for Deduction 3 (rewarded *Ss* will alternate responses and repeat stimuli in the Cross T maze).

REFERENCES

GLANZER, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45**, 387–393.

THOMPSON, M. E. Alternation in a T maze as a function of three variables. *Psychol. Rep.*, 1960, **7**, 103–110.

(Received November 4, 1960)

[1] This work was supported by Grant G-5817 from the National Science Foundation and was reported in a paper given at the 1960 meeting of the Rocky Mountain Psychological Association by the junior author.

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 630

SUPPLEMENTARY REPORT: UNLEARNING THE GAMBLER'S FALLACY[1]

HAROLD LINDMAN AND WARD EDWARDS

Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan

In a 1,000-trial probability learning experiment, Edwards (1961) found that as the length of a homogeneous outcome run increases, the probability that *S* will predict that event increases; this might be called a positive recency effect. Jarvik (1951) and Nicks (1959) have reported the opposite finding: as the length of a homogeneous outcome run increases, the probability that *S* will predict that event decreases. This is the gambler's fallacy. Jarvik's sequence had too few long outcome runs and so his *S*s' "gambler's fallacy" was not fallacious. But Nicks found a substantial gambler's fallacy even though he used a method of preparing sequences which makes every run statistic come out exactly at its expected frequency on the assumption of complete randomness and infinite sequence length. Edwards used the same sequence preparation procedure.

Edwards' (1961) 50–50 group data showed a small gambler's fallacy in the first 200 trials, no change of probability of prediction as a function of run length in the next 200 trials, and a positive recency effect in the last 600 trials. But even in the first 200 trials Edwards found less gambler's fallacy than did Nicks or Jarvik. Edwards and Nicks both used basic airmen at Lackland Air Force Base as *S*s. The disagreement between the Edwards and Nicks experiments indicates the desirability of a replication of the relevant part of Edwards' experiment using college student *S*s.

Method.—Twenty-five male undergraduates were tested in groups of 5 to 8 *S*s. All *S*s saw the same sequence. Two hundred stimulus cards, 100 red and 100 green, were presented in a random sequence prepared by the method of constrained randomization used by Edwards (1961) and Nicks (1959). Before each card was presented, *S* predicted whether the card would be red or green by writing R or G on a small card, which was then placed face down in a box. The *S*s were told to maximize correct predictions; no information about sequence was given except that the cards were "well mixed."

Results.—As shown in Fig. 1, the negative recency effect over the entire 200 trials was slightly greater than that found previously by Edwards (1961) but slightly less than that found by Nicks (1959). The tendency found by Edwards for the gambler's fallacy to be replaced by a positive recency effect was confirmed in this study, with the additional finding that this process is already under way by Trial 100.

It is interesting to note that Feldman (1959), using a sequence with more long runs than would be expected by chance and college student *S*s, found no gambler's fallacy at all in 200 trials. The gambler's fallacy apparently is a very evanescent phenomenon—unless it is more often right than wrong, as it was in Jarvik's (1951) experiment.

[1] This work was sponsored by the United States Air Force under Contract AF 49(638)-769, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

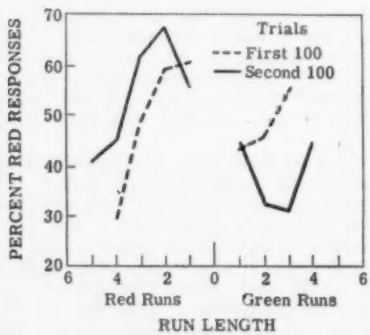


FIG. 1. Percentage of red responses following homogeneous outcome runs. (The *x* axis indicates the number of red or green outcomes included in the run. For example, the number 2 on the green runs side means that the number plotted is the percentage of times that a red response followed an outcome sequence consisting of a red followed by two greens. Runs which occurred less than three times per *S* in 100 trials are not plotted.)

REFERENCES

EDWARDS, W. Probability learning in 1,000 trials. *J. exp. Psychol.*, 1961, **62**, 381–390.

FELDMAN, J. On the negative recency hypothesis in the prediction of a series of binary symbols. *Amer. J. Psychol.*, 1959, **72**, 597–599.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

NICKS, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

(Received January 26, 1961)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 631

SUPPLEMENTARY REPORT: TWO-CHOICE DECISION BEHAVIOR WITH MANY ALTERNATIVE EVENTS

R. ALLEN GARDNER AND JOHN B. FORSYTHE

Wellesley College *Bureau of the Census*

With multiple-choice versions of the familiar probability learning experiment it has been demonstrated that $P(A)$, the relative frequency of choices of A, the most frequent alternative, is an increasing function of the number of alternatives (Cotton & Rechtschaffen, 1958; Gardner, 1957, 1958). The present study was an attempt to reproduce this result by varying the number of alternative events in the series, $A, B, C, \ldots, N$, while restricting S to two choices, A or not-A.

Method.—There were two parallel and independent experiments. Within each experiment, $\pi(A)$ was held constant and the number of alternative events was varied. In Exp. I, $\pi(A)$ was .70 and there were four conditions having either 2, 3, 4, or 7 alternative events. In Exp. II, $\pi(A)$ was .60 and there were four conditions having either 2, 3, 5, or 8 alternatives. The *S*s were 192 young soldiers drawn from the same population as those of two previous studies (Gardner, 1958, 1961). No S served in more than one condition and 24 served in each condition. The events consisted of block letters of the alphabet which were projected one at a time on the front wall of the experimental room. In all respects, the details of apparatus and procedure were identical to those of the two previous experiments with the following exception. Each response cabinet was provided with two telephone jacks and one plug. A small card was mounted over each jack. One card bore the letter designated as alternative A and the other was blank. The *S*s were instructed to predict on each trial whether or not the letter designated as alternative A would be presented next by inserting the plug into the appropriate jack, either the lettered jack or the blank jack.

Results.—The obtained values of $P(A)$ were tabulated in three periods of 140 trials: Period 1, Trials 1–140; Period 2, Trials 141–280; and Period 3, Trials 281–420. These are shown in Table 1. In all of the statistical tests quoted there, the score for each S was taken as the number of his choices of alternative A during the period in question. For convenience, the means have been presented in the table as proportions over 140 trials. The analysis of variance performed on these data was the same as that described in Gardner (1958, 1961). The *F*s for number of alternatives appear in the table and all are less than the 2.50 required for significance at the .05 level, most being less than 1.0. Moreover, the obtained values of $P(A)$ bear no orderly relationship to the number of alternatives in the series.

Under precisely the same conditions, but with the number of available choices kept equal to the number of alternative events in the sereis, $P(A)$ was a function of the number of alternatives (Gardner, 1958). In the present study, with S restricted to two choices, A or not-A, there was no detectable effect of the number of alternative events. Nevertheless, a variety of hypotheses remain open, for, when the number of choices was increased, and the number of alternative events held constant, the effect was distinctly different from that obtained when both the number of choices and alternative events were varied together (Gardner, 1961).

REFERENCES

COTTON, J. W., & RECHTSCHAFFEN, A. Replication Report: Two- and three-choice verbal-conditioning phenomena. *J. exp. Psychol.*, 1958, **56**, 96.

GARDNER, R. A. Probability-learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.

GARDNER, R. A. Multiple-choice decision-behavior. *Amer. J. Psychol.*, 1958, **71**, 710–717.

GARDNER, R. A. Multiple-choice decision-behavior with dummy choices. *Amer. J. Psychol.*, 1961, **74**, 205–214.

(Received January 24, 1961)

TABLE 1

MEAN PROPORTIONAL FREQUENCY ($P[A]$) OF CHOICE OF ALTERNATIVE A IN THE THREE SETS OF 140 TRIALS IN EXP. I AND II

Period	Trials	Alternatives in Exp. I, $\pi(A) = .70$					Alternatives in Exp. II, $\pi(A) = .60$				
		2	3	4	7	*F*	2	3	5	8	*F*
1	1–140	.621	.635	.640	.669	0.93	.563	.614	.580	.557	1.73
2	141–280	.721	.738*	.748*	.757*	0.61	.611	.643	.601	.607	0.63
3	281–420	.764*	.751*	.735	.762*	0.47	.633	.685*	.618	.659*	1.74

* By *t* test computed on raw scores, $P(A) > \pi(A)$, significant beyond the .05 level with 23 *df* (two-tailed test).

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 632

SUPPLEMENTARY REPORT: SHORT-TERM RETENTION AS A FUNCTION OF AVERAGE STORAGE LOAD[1]

KENNETH E. LLOYD

Washington State University

Lloyd, Reid, and Feallock (1960) and Reid, Lloyd, Brackett, and Hawkins (1961) have found recall errors to be systematically related to the average number of items *Ss* were required to remember. The present study supplements this work by extending the *average storage load* to higher values, and by using as items words judged to be unrelated.

The *Ss* listened to tape recorded sequences of items and recall points. An item consisted of a letter paired with a word; a recall point was indicated by the letter alone. The *Ss* were instructed to recall all words previously paired with a given letter at each recall point. Between a given item and its recall point in the sequence other items, other recall points, or both, could be interpolated. A sequence contained 6 subsequences which varied in average storage load from 2.5 to 12.5. Each subsequence contained 20 items and 10 recall points. At each recall point either 1, 2, or 3 items were requested. All subsequences with a given average storage load contained four individual loads which bracketed that average, e.g., a 4.5 subsequence contained individual loads of 3, 4, 5, and 6. Six sequences were recorded to partially counterbalance all conditions.

The 64 words were arbitrarily grouped into eight letter classes. Items were recorded at a 4-sec. rate, and recall points at a 3-sec. rate. Each of the 12 *Ss*, Washington State University undergraduates, served 5 days, and were reimbursed. They responded to two sequences on Days 1, 2, and 3, and to three sequences on Days 4 and 5.

[1] This research was supported by the United States Air Force through the United States Air Force Office of Scientific Research of ARDC under Contract No. AF 49(638)-805. This study is reported in greater detail in AFOSR Technical Note 60-1057, June 1960.

The means of errors over the repeated measures for each day were averaged for all *Ss*, and are shown in Table 1. While responding on Days 1 and 2, *Ss* looked at the word list to aid learning the arbitrary classes; this procedure facilitated recall. A trend analysis of transformed scores on Day 5 resulted in a significant overall trend ($F = 227.07$, $df = 5/55$, $P < .01$) as well as significance for the linear, quadratic, cubic, and quartic components (Fs = 288.77, 157.75, 76.31, and 12.00, respectively, df's = 1/11, Ps < .01). A second analysis of variance of raw scores for all 5 days yielded $P < .01$ for Load, Practice, and *Ss*, and for the L × S and P × S interactions. Load and Practice did not interact.

Ideally errors for each *S* should increase with increases in load. With repeated measures the 12 *Ss* provided 144 error curves. Inspection of adjacent curve points revealed 78 inversions out of 432 opportunities between the lower four load values. Between the higher three load values there were 139/288 inversions. This inversion probability of .48 suggests an asymptote at the higher loads.

In the three studies recall errors have increased regularly with average storage load over a wide range of its values, when related or unrelated items were used, and under conditions where *Ss* recalled one or several items at each recall point.

REFERENCES

LLOYD, K. E., REID, L. S., & FEALLOCK, J. B. Short-term retention as a function of the average number of items presented. *J. exp. Psychol.*, 1960, **60**, 201–207.

REID, L. S., LLOYD, K. E., BRACKETT, H. R., & HAWKINS, W. F. Short-term retention as a function of average storage load and average load reduction. *J. exp. Psychol.*, 1961, **62**, 518–522.

(Received January 25, 1961)

TABLE 1

MEAN ERRORS FOR AVERAGE STORAGE LOAD DURING 5 DAYS OF PRACTICE

Day	Average Storage Load											
	2.5		4.5		6.5		8.5		10.5		12.5	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	3.08	2.47	9.96	3.79	12.37	2.83	13.87	1.95	13.37	2.21	14.12	1.98
2	1.92	1.51	8.08	2.56	11.08	2.73	12.29	2.41	11.50	2.56	13.50	2.03
3	3.46	1.69	10.83	2.68	12.67	3.57	14.42	2.88	14.71	2.10	15.58	2.79
4	2.37	1.52	8.97	2.15	12.31	2.58	13.97	2.88	13.62	1.82	14.37	2.62
5	1.52	1.46	8.12	2.92	10.61	2.65	12.01	2.38	12.26	2.47	13.70	2.41

ZEITSCHRIFT FÜR EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland
PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Herausgeber für Österreich
PROF. DR. H. ROHRACHER, Wien

Herausgeber für die Schweiz
PROF. DR. H. MEILI, Bern

Redaktion
DR. C. J. HOGREFE, Göttingen

Band VIII/2

RÜDIGER SEIFERT, **Untersuchungen zur Korrelation psychischer Symptome mit den Wirkungen einer Belastung durch Sauerstoffmangel**

RUDOLF SCHAEDELI, **Untersuchungen zur Verifikation von Meilis Intelligenzfaktoren**

MERVYN WILLIAM PERRINE, **Die Bedeutung des Handelns für das Wahrnehmungslernen**

MARIANNE FISCHER UND EBERHARD ULICH, **Über die Abhängigkeit einer kurzzeitigen Konzentrationsleistung von der Tageszeit bei Kindern und Jugendlichen verschiedenen Alters**

LILIAN BLÖSCHL, **Neue Experimente zum Problem der vermittelten Assoziationen**

WERNER H. TACK, **Die Beschreibung diskreter Verteilungen durch informationstheoretische Maße**

THEO HERRMANN, **Erwiderung zu W. H. Tack: Die Beschreibung diskreter Verteilung durch informationstheoretische Maße**

Kritische Buchbesprechung
RAYMOND B. CATTELL, **Personality and Motivation Structure and Measurement** (HANS HÖRMANN)

All articles are provided with summaries in English and French.
The journal appears quarterly and contains 160 pages per number.

Subscription is DM 64.—per Volume

VERLAG FÜR PSYCHOLOGIE—DR. C. J. HOGREFE—GÖTTINGEN
(Brentanoweg 10, Germany)

Effective *January 1, 1962*, **all regular** articles, **supplementary reports, and replication reports** should be submitted to the Editor-elect: David A. Grant; Department of Psychology, University of Wisconsin; Madison 6, Wisconsin. **Early publication** articles should be sent to the Editor: Arthur W. Melton; Department of Psychology, University of Michigan; Ann Arbor, Michigan until *July 1, 1962*. Thereafter they should be sent to the Editor-elect.

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.